Assessment of performance in the presence of undesirable outputs: the promotion of livability and sustainable development of cities and countries using Data Envelopment Analysis

Andreia Zanella

A thesis submitted to Faculdade de Engenharia da Universidade do Porto
for the doctoral degree in Industrial Engineering and Management

Supervisors

Professor Ana Maria Cunha Ribeiro dos Santos Ponces Camanho
Professor Maria Teresa Galvão Dias

U.PORTO
FEUP FACULDADE DE ENGENHARIA
UNIVERSIDADE DO PORTO

2014

Abstract

This thesis is concerned with the development of innovative models for the assessment and monitoring of performance using Data Envelopment Analysis (DEA) in the presence of undesirable outputs. The models focus on the construction of composite indicators and are applied to the evaluation of cities and countries with the aim to promote livability and sustainable development. The empirical part of the thesis contributes to the definition of better public policies through the identification of best practice examples and areas with more potential for improvements.

The thesis includes four main research topics. The first topic discusses two alternative approaches to incorporate undesirable outputs in composite indicators constructed using DEA models. The first is an indirect approach, based on a standard DEA model which includes a transformation in the measurement scale of the undesirable outputs. The second is a direct approach, based on a DEA model specified with a directional distance function. This topic also discusses the incorporation of restrictions to weights in this context, and proposes a novel specification of assurance region type I weight restrictions.

The second topic approaches the measurement of productivity change over time in the presence of undesirable outputs using the Malmquist-Luenberger index and the Luenberger index. An enhanced version of the Malmquist-Luenberger index is proposed. The results obtained using the different productivity indices are compared and discussed.

The third topic assesses the environmental performance of countries worldwide. The assessment is conducted using the composite indicator specified using the indirect approach proposed in chapter 3. It enables benchmarking in such a way that it is possible to identify the strengths and weaknesses of each country, as well as the peers with similar features to the country under assessment.

The last topic develops a framework to assess the livability of European cities covering two components: human well-being and environmental impact. It is proposed a conceptual model that extends the concept of urban livability to include a component related to environmental sustainability. Then, the measurement of cities' livability is conducted using the composite indicator specified with a directional distance function, as developed in chapter 3. Finally, the evolution of cities' performance over time is assessed using the Luenberger productivity index.

Overall, this thesis contributed to the development of robust tools to evaluate and promote livability and sustainable development in cities and countries, with a view to foster better standards of living now and in the future.

Resumo

Esta tese concentra-se no desenvolvimento de modelos inovadores para a avaliação e monitorização de desempenho utilizando a técnica *Data Envelopment Analysis* (DEA) na presença de *outputs* indesejáveis. Os modelos desenvolvidos visam a construção de indicadores compósitos e são aplicados na avaliação de cidades e países com o objetivo de promover o bem-estar e o desenvolvimento sustentável. A parte empírica da tese contribui para a definição de melhores políticas públicas através da identificação dos exemplos de boas práticas e das áreas com maior potencial de melhoria.

Esta tese divide-se em quatro tópicos principais. O primeiro tópico discute duas abordagens alternativas que podem ser usadas para incorporar *outputs* indesejáveis em indicadores compósitos construídos com base em modelos de DEA. A primeira é uma abordagem indireta, baseada num modelo tradicional de DEA, que inclui uma transformação na escala de medição dos indicadores indesejáveis. A segunda é uma abordagem direta, baseada num modelo especificado com uma função de distância direcional. Esse tópico também discute a incorporação de restrições de pesos, e propõe uma nova especificação de restrições de pesos do tipo *assurance region type I*.

O segundo tópico aborda a análise da evolução da produtividade ao longo do tempo na presença de *outputs* indesejáveis utilizando os índices de *Malmquist-Luenberger* e *Luenberger*. Neste tópico também se propõe uma versão melhorada do índice *Malmquist-Luenberger*. Os resultados obtidos utilizando os diferentes índices de produtividade são comparados e discutidos.

O terceiro tópico aborda a avaliação do desempenho ambiental de países. A avaliação é conduzida por meio do indicador compósito proposto no capítulo 3, baseado na abordagem indireta. Este indicador compósito permite identificar as forças e fraquezas de cada país, bem como os países que possuem características similares às do país em avaliação e que podem ser considerados exemplos de boas práticas.

O último tópico abordado nesta tese desenvolve uma ferramenta para avaliar a habitabilidade das cidades europeias englobando duas componentes: bem-estar humano e impacto ambiental. Propõe-se um modelo conceptual que alarga o conceito de habitabilidade urbana de forma a incluir uma componente relacionada com a sustentabilidade ambiental. Tendo por base o modelo conceptual proposto, a habitabilidade das cidades é então avaliada utilizando o indicador compósito desenvolvido no capítulo 3, especificado com uma função de distância direcional. Por fim, a evolução do desempenho das cidades é avaliada utilizando o índice de *Luenberger*.

Em síntese, esta tese contribui para o desenvolvimento de ferramentas robustas para avaliar e promover o bem-estar e o desenvolvimento sustentável

em cidades e países, com vista a melhorar os padrões de vida nos dias atuais e no futuro.

Acknowledgments

I would like to thank to my supervisors, Professor Ana Camanho, for her excellent guidance, constant encouragement and for sharing her knowledge, which were crucial to the completion of this thesis, and Professor Maria Teresa Galvão Dias for her comments, encouragement and support during the doctoral research.

I am grateful to the Doctoral Program in Industrial Engineering and Management for welcoming me and making this research work possible.

I thank the financial support of the Portuguese Foundation for Science and Technology (FCT) through the projects iTeam (Pt/SES-SUES/0041/2008) and Mesur (PTDC/SEN-ENR/111710/2009).

Also, I would like to thank my colleagues and friends, especially Isabel and Vera for sharing experiences and knowledge, and Carla and Maria for their friendship and support during these four years.

I am grateful to my family for all the kindness and understanding over these years I have been living far away from them. I thank particularly my mother, for the support, encouragement and confidence placed in me. I certainly owe her the important steps I took in my life.

Finally, I would like to thank Fabrício for his patience, help, and love throughout the time of this research and especially for encouraging me to come to Portugal to do this course.

Acronyms

ARI Assurance Region type I

ARII Assurance Region type II

CI Composite Indicator

CCPI Climate Change Performance Index

CRS Constant Returns to Scale

DEA Data Envelopment Analysis

DRS Decreasing Returns to Scale

DMU Decision Making Unit

DDF Directional Distance Function

EC Efficiency Change

EPI Environmental Performance Index

EU European Union

GDP Gross Domestic Product

IRS Increasing Returns to Scale

L Luenberger

MI Malmquist Index

ML Malmquist-Luenberger

NDRS Non-Decreasing Returns to Scale

NIRS Non-Increasing Returns to Scale

OECD Organisation for Economic Co-operation and Development

PPS Production Possibility Set

SFA Stochastic Frontier Analysis

TC Technical Change

VRS Variable Returns to Scale

Table of Contents

List of Figures

List of Tables

Chapter 1

Introduction

1.1 General context

The assessment and promotion of urban livability and sustainable development is an issue with growing importance among scientific and policy-making communities. Efforts from academy and governmental institutions are leading to a better understanding of how local communities, cities, and countries are performing compared to their peers, and encouraging the monitoring of progress over time.

Urban livability is nowadays recognized as an important component of competitive advantage. It can be defined as the suitability of a given place for human living (Merriam-Webster, 2013). In the past few years, indicators that can measure urban livability have been developed to guide decisions about where to invest in a new business or where to seek employment (Australian Department of Infrastructure and Transport, 2012). National and local authorities are increasingly supporting efforts to improve the understanding of cities' progress in terms of productivity, sustainability and livability, such as the yearly report "State of Australian Cities" by the Australian Government (Australian Department of Infrastructure and Transport, 2012).

In addition to promoting competitive advantage by improving urban livability, governments can answer claims of more exigent citizens demanding better quality of life, facilities and infrastructure. Governments also face pressures from international agreements, such as the Kyoto Protocol (United Nations, 1998) or the European Union climate and energy package (European Union, 2008), environmentalists, and society to control and reduce emissions and preserve the natural resources and environment. Therefore, the governments' challenge lies in providing good standards for human living but with a view to preserving the natural resources and environmental conditions. This requires the definition of economic and social development plans to comply with the sustainable development concept.

Since the late 80's, the definition of sustainable development proposed by the World Commission on Environment and Development is widely accepted. It states that sustainable development implies "meeting the needs of the present without compromising the ability of future generations to meet their own needs" (United Nations, 1987). Sustainable development is often presented as the integration of three interdependent components: economic, social and environmental. In the past few years, efforts to monitor these components have generated a large number of indicators intended to measure the performance in aspects related to emissions, waste production, green space, safety, income, health and education, among others.

The majority of countries and cities collects data on these indicators, but the amount of data generated is often too large and it is not sufficiently clear to provide useful information and practical guidance to attend the policy-makers needs. So, many of these indicators have been aggregated to build composite indicators (CIs).

A composite indicator is given by the aggregation of several individual indicators in a single measure. It has benefits such as the capacity to summarize information, the facility to interpret results compared with a battery of separate indicators, and the capacity to reduce the visible size of a set of indicators without dropping the underlying base information (Nardo et al., 2008). Examples of well-established composite indicators are the Environmental Performance Index (Emerson et al., 2012), Climate Change Performance Index (Burck et al., 2012), and the Human Development Index (United Nations, 2013). Although a considerable amount of research related to the construction of composite indicators has been developed in recent years (as reviewed in Nardo et al. (2008)), they are not effective in providing managerial information to guide improvements. Furthermore, they are prone to criticism regarding the subjectivity inherent in the specification of the relative importance given to the individual indicators in the construction of the CI.

The Organisation for Economic Co-operation and Development (OECD) and the European commission provide a handbook for the construction of composite indicators that discusses the range of methodological approaches available to construct CIs (Nardo et al., 2008). The handbook highlights the growing interest in composite indicators by the academic circles, media and policymakers. One point discussed and recognized as a source of contention is the definition of the relative importance of the indicators. The handbook points Data Envelopment Analysis (DEA) as an interesting weighting and aggregation procedure to reduce the inherent subjectivity associated with the specification of weights. As the indicator weights result from an optimizing process based on linear programming, they are less prone to subjectivity and controversy.

The broad subject area of this thesis is the use of frontier analysis methods, in particular DEA, for the assessment and promotion of urban livability and sustainable development. The motivation and main objectives of this thesis are discussed in detail in the next section.

1.2 Motivation and research objectives

Despite the growing interest and work in the field of performance assessment, robust tools to measure and guide improvements towards more livable and sustainable cities and countries are not available. More sophisticated and robust methods for performance assessment and benchmarking in this context are worth exploring.

This research contributes to improve the current methods of livability and sustainability assessment by exploring new methodologies, based on frontier techniques, that can effectively provide enhanced managerial information and guide cities and countries towards sustainable development.

Frontier methods have the advantage of allowing comparisons with the best observed performance by constructing a best practice frontier based on empirical data. From the alternative frontier methods available, in this thesis it was chosen to explore in detail the use of Data Envelopment Analysis due to the greater flexibility to incorporate the multidimensional nature of the livability and sustainable development concepts, and the use of minimal assumptions on the shape of the best practice frontier.

The DEA technique uses linear programming to evaluate the relative efficiency of an homogeneous set of decision making units (DMUs) in their use of multiple inputs to produce multiple outputs. In addition to providing a single overall measure of performance and being able to fight the subjectivity associated with the specification of the indicators weights, the performance assessment using DEA allows the identification of areas for improvement and best practice examples. These properties make this technique particularly valuable for conducting benchmarking in the context of cities and countries livability and sustainable development. The advantages of the DEA technique that motivate its use in this thesis are described in greater detail in chapter 2.

Standard DEA models assume that the individual output indicators represent good aspects, so they are measured on a scale for which higher values correspond to better performance. However, in real-world applications, both desirable and undesirable outputs indicators may be present. For example, in environmental performance assessment we may have an output indicator related to quality of the water, for which more output corresponds to better performance and another output indicator related to the levels of CO_2 emissions, for which less output corresponds to better performance. In this situation, an inefficient DMU should increase the quality of the water and/or decrease the levels of CO_2 emissions to improve performance. Although the literature addresses the construction of DEA models with undesirable outputs, this issue is not discussed in the context of evaluations using composite indicators. Handling undesirable outputs in performance assessments using DEA requires particular attention as the alternative treatments may have a huge impact on the results.

This thesis has two main objectives. The first is to develop innovative models for the assessment of performance in the presence of undesirable outputs using DEA. These models will be focused on applications involving the aggregation of key performance indicators. The fulfilment of this first objective opens the possibility to accomplish the second objective, that consists in undertaking a comprehensive evaluation of cities and countries, aiming to promote urban livability and sustainable development. In addition to assessing performance and monitoring its evolution over time, the models developed in this thesis can be used to identify best practice examples and areas in which cities and countries have more potential for improvements. Although the models are applied to the assessment of urban livability and sustainable development, they are easily transposable to different contexts and replicable over time.

Given these main objectives, the specific objectives of the research described in this thesis are as follows:

1. To review the current approaches available to treat undesirable outputs in DEA models and develop a DEA-based composite indicator for the evaluation of performance in the presence of undesirable outputs;

2. To incorporate in the composite indicator information on the relative importance of individual indicators, in percentual terms, using weight restrictions;

3. To explore the different DEA-based approaches that can be used to accommodate undesirable outputs in the analysis of productivity change over time, namely the ratio-based Malmquist-Luenberger index and the difference-based Luenberger index;

4. To assess countries environmental performance using DEA, based on the aggregation of the indicators that underlie the estimation of the Environmental Performance Index (EPI), identifying the factors corresponding to the best and worst features of each country, as well as the peers with similar features to the countries with worse performance;

5. To define an appropriate set of indicators to assess cities' livability extending the concept of urban livability to include a component related to environmental sustainability;

6. To assess the livability of European cities and provide managerial information for performance improvement. The managerial information is delivered through the identification of peers whose practices are examples to be followed and the identification of the areas in which each city has the best and worst features;

7. To analyse the cities evolution over time in terms of livability using the Luenberger productivity index, and identify the innovative cities, i.e. those

cities that are responsible for movements of the production frontier towards better productivity levels.

1.3 Thesis summary

This thesis is structured in seven chapters, which are briefly described in this section.

Chapter 2 presents an introduction to the methods for the evaluation of performance using frontier techniques. Particular emphasis is given to the DEA method, which is the core method used for the achievement of the research objectives stated for this thesis.

Chapters 3 and 4 present the theoretical developments of the thesis. Each chapter includes a detailed literature review related to the topic approached.

Chapter 3 is organized in three parts. The first reviews and discusses the literature related to the incorporation of undesirable outputs in DEA. The second addresses the incorporation of undesirable outputs in the context of the construction of composite indicators using DEA. Two different approaches are discussed and compared: an indirect approach, based on a standard DEA model that includes a transformation in the measurement scale of the undesirable outputs, and a direct approach, based on a DEA model specified using a directional distance function. Finally, the third part discusses the incorporation of restrictions to weights in the context of assessments involving composite indicators. Restrictions to weights can be included in the model in order to reflect decision-maker preferences about the relative importance of individual indicators, or/and to improve the discrimination among the DMUs' performance scores. In both cases, the restrictions to weights prevent DMUs from obtaining a high score only owing to a judicious choice of weights.

The first part of chapter 4 discusses the different approaches that can be used to accommodate undesirable outputs in the analysis of productivity change over time. In particular, the ratio-based Malmquist-Luenberger (ML) index, derived from a standard output oriented Malmquist index, and the difference-based Luenberger productivity index are reviewed. In the second part of the chapter, it is shown that an alternative Malmquist-Luenberger index can be derived from a standard input oriented Malmquist index. The two versions of the ML index (input and output oriented versions) represent equally good adaptations of the Malmquist index, although they give different results. In order to avoid an arbitrary selection of the input or output ML indices, we propose the use of an enhanced ML index. An empirical example is used to compare the results obtained by the different versions of the ML indices with the results of the Luenberger index, which is a well-established measure of productivity change over time considering simultaneous adjustments to inputs as well as desirable and undesirable outputs.

Chapters 5 and 6 correspond to empirical applications. Chapter 5 illustrates the application of one of the models discussed in chapter 3, corresponding to the indirect approach to treat undesirable outputs, whereas chapter 6 illustrates the application of the alternative approach, corresponding to the direct formulation to treat undesirable outputs in composite indicators. Next, each chapter is explained in more detail.

Chapter 5 uses an enhanced DEA model that includes a transformation to the measurement scale of undesirable outputs, to provide a single summary measure of countries environmental performance. This assessment is based on the aggregation of the indicators that underlie the estimation of the Environmental Performance Index (EPI). This model enables benchmarking in such a way that it becomes possible to identify the strengths and weaknesses of each country, as well as the peers with similar features to the country under assessment.

Chapter 6 develops a tool to assess livability of European cities covering two components: human wellbeing and environmental impact. First, it is proposed a conceptual model to assess cities' livability, that extends the concept of urban livability to include a component related to environmental sustainability. Then, the measurement of cities' livability is conducted using the CI based on the directional distance function model developed in chapter 3. One of the innovative features of the model proposed is to allow the specification of different directional vectors that can focus on specific components of livability (e.g., human wellbeing or environmental impact). Finally, it is also assessed the evolution of cities' performance over time using the Luenberger productivity index.

Chapter 7 presents the conclusions, main contributions of the thesis and suggestions for future research.

Chapter 2

The assessment of performance

2.1 Introduction

This chapter provides the foundations of the performance measurement methods that will be used throughout the thesis. The approaches to evaluate performance using frontier techniques were first developed in the 1970s. Frontier techniques estimate the maximum level of output that a producer can obtain for a given level of resources, or the minimum level of resources that is required to achieve a given level of outputs.

These approaches to evaluate performance followed two distinct routes that differ in the way the frontier is estimated, corresponding to a parametric (econometric) or a nonparametric (mathematical programming) approach. This chapter provides an introduction to performance measurement using frontier techniques, focusing on the nonparametric approach, as the models developed in this thesis followed this route.

Particular emphasis is given to Data Envelopment Analysis (DEA), as it is the core method used throughout the thesis. The theory of production underlying the DEA method relies on the axiomatic approach of Shephard (1970), which is based on productions sets.

This chapter is organised in two main sections. First, an introduction to the theory of production is presented, including the main concepts and a brief historical overview of the evolution of this scientific area. Some important definitions concerning the axiomatic approach, production sets and relative efficiency are provided. Then, the Data Envelopment Analysis (DEA) technique is described in more detail. It is provided a description of the theory underlying the construction of the DEA frontier and the measurement of efficiency.

2.2 Production theory

2.2.1 Production processes

Production can be viewed as a process in which a firm, often called decision making unit (DMU), uses inputs to produce outputs, as shown schematically in Figure 2.1. For example, a factory may use as inputs raw materials, labour and capital in order to produce goods. Any other process taking inputs to produce outputs can be viewed in this way.

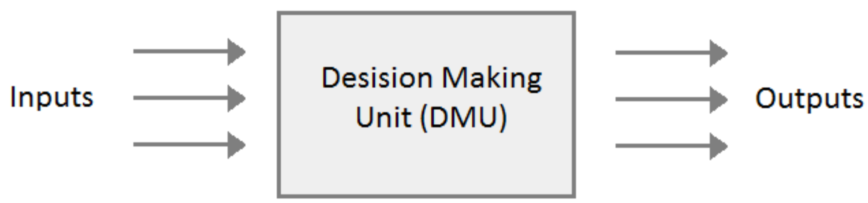


Figure 2.1: The production process

In performance assessment, the selection of the inputs and outputs is a critical part of the process. It is clear that the results obtained are strongly dependent on the inputs and outputs selected, thus they should be carefully chosen in order to reflect the primary aims of the assessment. Furthermore, the amount of output produced must be related in some way to the amount of inputs used to secure it. This relation is called technology of production and it defines a conversion process in which the outputs are intended to be maximised subject to a set of inputs, or alternatively the inputs are required to be minimized subject to a set of outputs.

Exact knowledge of the technology of production is not often available, and several different methods have been developed to represent it. These methods generally consist of empirical estimations of the location of a frontier that envelops the DMUs observed. Regardless of the differences in the methods available to estimate the technology of production, the specification of a best practice reference, corresponding to the frontier of the production possibility set, enables to quantify a measure of efficiency for each DMU, involving a comparison between the observed values of the outputs and inputs of the DMU under assessment and the optimal values corresponding to a point on the frontier.

2.2.2 Production sets and the axioms of production

Consider a production process involving n DMUs, $j = 1, ..., n$, consuming m inputs x_{ij}, $i = 1, ..., m$, to produce s outputs y_{rj}, $r = 1, ..., s$. The technology, or production possibility set (PPS), can be specified as shown in (2.1).

$$T = \{(x, y) : x \text{ can produce } y\} \tag{2.1}$$

The production possibility set T contains all feasible input-output combinations corresponding to a certain production process.

The postulates for the construction of the production possibility set T can be defined as follows (Banker (1984); Banker and Thrall (1992); Fare and Grosskopf (2005)):

Postulate 1. Inclusion of observations.

All observed DMUs are included in the production possibility set, i.e., $(x_j, y_j) \in T$, for all $j = 1, ..., n$.

Postulate 2. No outputs can be produced without some input.

Zero output can be produced by any input vector $x \in \Re^m_+$, but it is impossible to produce output without any inputs, i.e. if $y \geq 0$, $y \neq 0$, $x = 0$ then $(x, y) \notin T$.

Postulate 3. Inefficiency.

(a) If $(x, y) \in T$ and $x' \geq x$, then $(x', y) \in T$.
(b) If $(x, y) \in T$ and $y' \leq y$, then $(x, y') \in T$.

Postulate 4. Ray unboundedness.

If $(x, y) \in T$ then $(\lambda x, \lambda y) \in T$, for all $\lambda > 0$.

Postulate 5. Closedness: Technology T is a closed set.[1]

Given a closed correspondence, represented by $\to$, if $(x_j, y_j) \to (x', y')$ and $(x_j, y_j) \in T$ for all $j = 1, ..., n$, then $(x', y') \in T$.

Postulate 6. Convexity.

If $(x_j, y_j) \in T$, $j = 1, ..., n$, and λ_j are nonnegative scalars such that $\sum_{j=1}^{n} \lambda_j = 1$, then $\left(\sum_{j=1}^{n} \lambda_j \ x_j, \sum_{j=1}^{n} \lambda_j \ y_j\right) \in T$.

[1]Given the imposition of the ray unboundedness postulate, implying the existence of constant returns to scale, the closedness postulate is not required, as the technology is always closed (Podinovski and Bouzdine-Chameeva, 2011). For other variants of the postulates defining the production possibility set, involving changes to the ray unboundedness postulate, the closedness postulate needs to be explicitly stated.

Postulate 7. Minimum extrapolation.

T is the intersection of all sets $\hat{T}$ that satisfies the postulates 1, 2, 3, 4, 5 and 6.

The previous postulates allow defining a Constant Returns to Scale (CRS) production possibility set. Dropping the ray unboundedness postulate will lead to the definition of a Variable Returns to Scale (VRS) production possibility set. The returns to scale is a characteristic of the boundary of the production technology. It is used to define how the technology behaves when changes to the scale of operation occur.

To explain this concept, consider the case of a DMU that operates on the boundary of the PPS and uses a single input x to produce a single output y. If a change in the scale of operation occurs and the output increases proportionally to the input, it is said that the boundary at that point exhibits Constant Returns to Scale (CRS). If output increases proportionally more than the input, the boundary at that point exhibits Increasing Returns to Scale (IRS), and if the output increases proportionally less than the input, the boundary at that point exhibits Decreasing Returns to Scale (DRS). The term Variable Returns to Scale (VRS) is used to denote a boundary that exhibits any combination of IRS and/or DRS with CRS.

The production technology T can alternatively be represented by input or output sets. The input requirement set, $L(y)$, describes the set of all input vectors x which can be used to produce the output vector y, as shown in (2.2).

$$L(y) = \{x : y \text{ can be produced by } x\} = \{x : (x, y) \in T(x, y)\} \tag{2.2}$$

The input set is bounded from the input isoquant which contains the minimum inputs necessary to secure a certain output. The mathematical formulation of the input isoquant is defined as shown in (2.3).

$$IsoqL(y) = \{x : x \in L(y), \delta x \notin L(y) \text{ for } \delta < 1\} \tag{2.3}$$

The isoquant defines a frontier to the input set. Those DMUs that lie on the frontier are efficient in the sense that radial (proportional) reduction of inputs is not possible.

For the example shown in Figure 2.2, involving two inputs and a single output, the input set $L(y)$ contains all the input vectors within the isoquant $IsoqL(y)$. This isoquant is formed by the segments linking the points B, C and D, and the rays parallel to the axes, i.e., the ray starting in B and passing through A, and the ray starting in D and passing through E. Note that in the rays $\overrightarrow{BA}$ and $\overrightarrow{DE}$, the proportional reduction of both inputs is not possible, although it is possible to reduce only one input by moving along the rays to reach the efficient frontier at B or D.

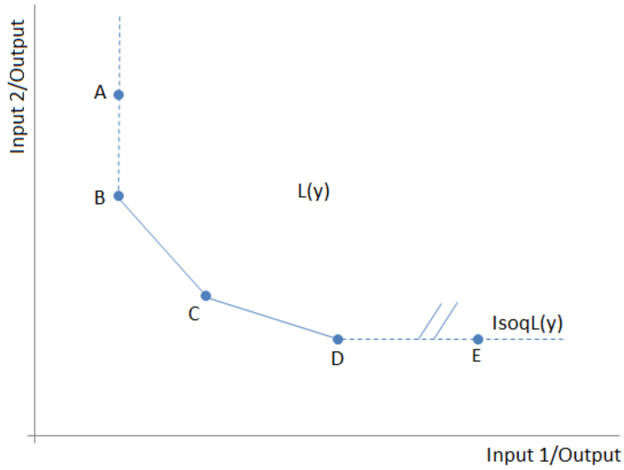


Figure 2.2: The input set

Similarly, the production technology can be represented by an output set

$P(x)$. It describes the set of all output vectors y which can be produced using the input vector x, as shown in (2.4).

$$P(x) = \{y : x \text{ can produce } \ y\} = \{y : (x, y) \in T(x, y)\} \tag{2.4}$$

The output set is bounded by the output isoquant that contains all the output combinations that cannot be proportionally increased without increasing the input vector x. It is defined as shown in (2.5).

$$IsoqP(x) = \{y : y \in P(x), \theta y \notin P(x) \text{ for } \theta > 1\} \tag{2.5}$$

The concepts underlying the representation of the technology for the output set are illustrated in Figure 2.3, using two outputs that are produced with a single input.

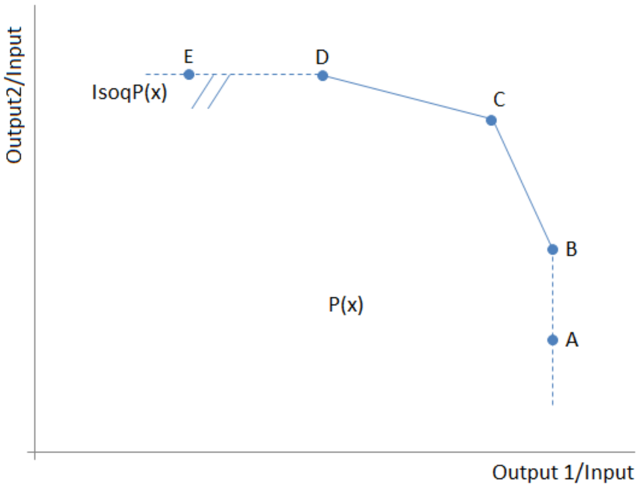


Figure 2.3: The output set

The output set $P(x)$ is formed by all points in the production possibility set bounded by the isoquant, i.e., by the segments linking points B, C and

D and the rays parallel to the axes, spanned from points B and D. The points that lie on the rays $\overrightarrow{BA}$ and $\overrightarrow{DE}$ cannot be proportionally increased keeping the current input levels, as only one of the outputs can be increased, i.e., output 2 when moving along $\overrightarrow{BA}$ in direction to B, or output 1 when moving along $\overrightarrow{DE}$ in direction to D.[2]

2.2.3 Technical efficiency

Efficiency involves a comparison of the actual location of a DMU within the PPS with the optimal input and output levels corresponding to the points located on the production frontier. The first measure of technical efficiency dates back to the work of Debreu (1951) and Farrell (1957). The Debreu's measure of efficiency was called the "coefficient of resource utilisation". Farrell extended this previous work by proposing the measurement of efficiency using empirical observations, i.e. by comparing a DMU to the best actually achieved by its peers.

The DMUs' efficiency measure can be obtained from two perspectives, corresponding to an input-reduction or output-expansion orientation. Assuming an input orientation, the Debreu-Farrell technical efficiency measure is defined as the maximum radial (proportional) reduction to all inputs that is feasible to achieve whilst securing a certain output level, within a given technology. On the other hand, assuming an output orientation, this measure is defined as the maximum radial (proportional) expansion to all outputs that is feasible to achieve with a certain input level, within a given technology (Fried et al., 2008).

In order to relate the Debreu-Farrell measures to the structure of the pro-

[2]The production technologies represented in Figures 2.2 and 2.3 are piecewise linear. While the econometric approach estimates smooth parametric frontiers, the mathematical programming approach estimates piecewise linear frontiers. As this thesis will use a mathematical programming approach, we will present the illustration of piecewise linear frontiers only.

duction technology, consider the technologies defined in (2.2) and (2.4).

The Debreu-Farrel input oriented measure of technical efficiency (TE_i) can be formally defined as the value of the function shown in (2.6).

$$TE_i(x, y) = \min\{\delta : \delta x \in L(y)\} \tag{2.6}$$

Note that, for $x \in L(y)$, $\delta \leq 1$, and for $x \in IsoqL(y)$, $\delta = 1$.

Figure 2.4 graphically illustrates the input oriented technical efficiency (TE_i) measure using a sample of seven DMUs. DMUs A, B, C, D and E lie on the frontier defined by the isoquant $IsoqL(y)$, and thus are considered efficient in the Debreu-Farrell sense ($\delta = 1$). For DMUs F and G, the technical efficiency measure is given by the ratios $\frac{OF^*}{OF}$ and $\frac{OG^*}{OG}$, respectively, and is less than one for both DMUs.

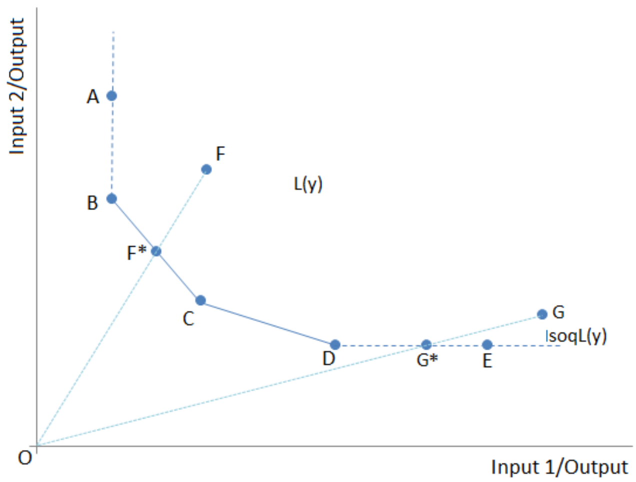


Figure 2.4: Technical efficiency measure for the input set

Similarly, the Debreu-Farrell output oriented measure of technical efficiency can be defined as the value of the function shown in (2.7).

$$TE_o(x, y) = \max\{\theta : \theta y \in P(x)\} \tag{2.7}$$

Again, note that for $y \in P(x)$, $\theta \geq 1$, and for $y \in IsoqP(x)$, $\theta = 1$.

In Figure 2.5, the output oriented technical efficiency measure is illustrated for a set of seven DMUs. The DMUs A, B, C, D and E are considered efficient in the Debreu-Farrell sense as they lie on the isoquant $IsoqP(x)$, so $\theta = 1$. The technical efficiency measure for DMUs F and G is given by the ratios $\frac{OF^*}{OF}$ and $\frac{OG^*}{OG}$, respectively, and its value is greater than one for both DMUs. [3]

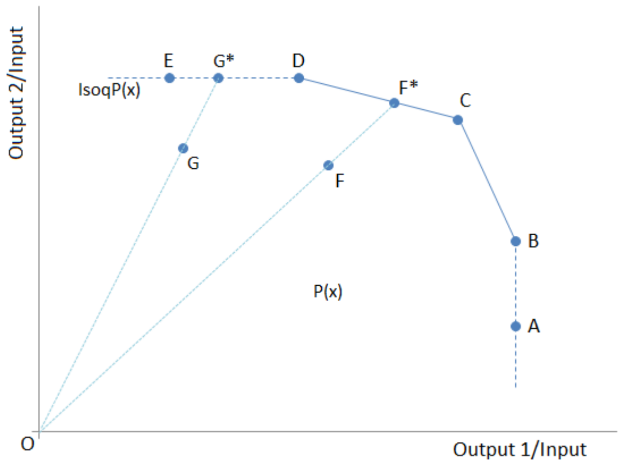


Figure 2.5: Technical efficiency measure for the output set

However, the Debreu-Farrell definition of efficiency, also known as radial efficiency, is not sufficient for defining the "truly" efficient DMUs. A stronger notion of efficiency was provided by Koopmans (1951). The author stated that "a producer is technically efficient if an increase in any output requires a reduction in at least one other output or an increase in at least one input,

[3]This interpretation follows the convention of defining efficiency as a ratio of optimal to actual. Fried et al. (2008) noted that some authors replace (2.7) with $TE_o(x, y) = [\max\{\theta : \theta y \in P(x)\}]^{-1}$.

and if a reduction in any input requires an increase in at least one other input or a reduction in at least one output".

The Debreu-Farrell's efficiency is based on the radial contraction (expansion) of inputs (outputs). This implies that after the equiproportional reduction (expansion) of the inputs (outputs) of a DMU leading to a point on the boundary of the PPS, there is no scope for further improvement for at least one input (output). On the other hand, Koopmans' efficiency investigates further the potential reduction (expansion) of each input (output) beyond the radial movements.

Considering the technologies defined in (2.2) and (2.4), the Koopmans' notion of efficiency can be represented by the input and output efficient subsets, given by the expressions shown in (2.8) and (2.9), respectively.

$$E(y) = \{x : x \in L(y), x' \leq x \text{ and } x' \neq x \Rightarrow x' \notin L(y)\} \tag{2.8}$$

$$E(x) = \{y : y \in P(x), y' \leq y \text{ and } y' \neq y \Rightarrow y' \notin P(x)\} \tag{2.9}$$

Therefore, $E(y)$ and $E(x)$ are subsets of the isoquants $IsoqL(y)$ and $IsoqP(x)$, respectively. The efficient subsets $E(y)$ and $E(x)$ differ from the isoquants as the latter may contain rays parallel to the axes, which are not part of the efficient frontier in Koopmans' sense.

These concepts are illustrated in Figures 2.4 and 2.5. In both figures the points A, E and G^* satisfy the Debreu-Farrell conditions for technical efficiency but they do not satisfy the Koopmans' conditions. For point G^*, for example, in the case of the input oriented approach (Figure 2.4), it is possible to reduce the consumption of input 1, without increasing input 2 and keeping the current output level, i.e. to move point G^* horizontally until reaching point D. Similarly, in the case of the output oriented approach (Figure 2.5), it is possible to increase the production of output 1, keeping the current input level and without reducing output 2, i.e. to move point G^* horizontally until reaching point D. Conversely, points B, C, D and F^* satisfy both definitions of technical efficiency.

In the remainder of this thesis, the concept of technical efficiency adopted will follow the Koopmans' criteria, as it is consensual in the literature that it corresponds to a stronger notion of efficiency, and thus should underlie the analysis of performance based on quantitative methods.

2.2.4 Parametric and nonparametric frontiers

The performance measurement methods that rely on the estimation of a frontier have evolved following two parallel research lines: a parametric and a nonparametric approach. These methods differ in the way the frontier is specified and estimated.

The parametric approach specifies the frontier as a function with a precise mathematical form (usually the translog or the Cobb-Douglas functions). It requires an *a priori* specification of the functional form representing the frontier. On the other hand, the nonparametric approach does not require any assumptions with respect to the functional form of the frontier. The frontier is defined by a set of postulates that the points on the boundary of the production possibility set have to satisfy.

Figure 2.6 classifies the various types of parametric and nonparametric frontiers. The most common methods for efficiency evaluation are Data Envelopment Analysis (DEA) in the nonparametric literature, and Stochastic Frontier Analysis (SFA) in the parametric literature. Both the parametric and nonparametric methods can be further divided into stochastic and deterministic.

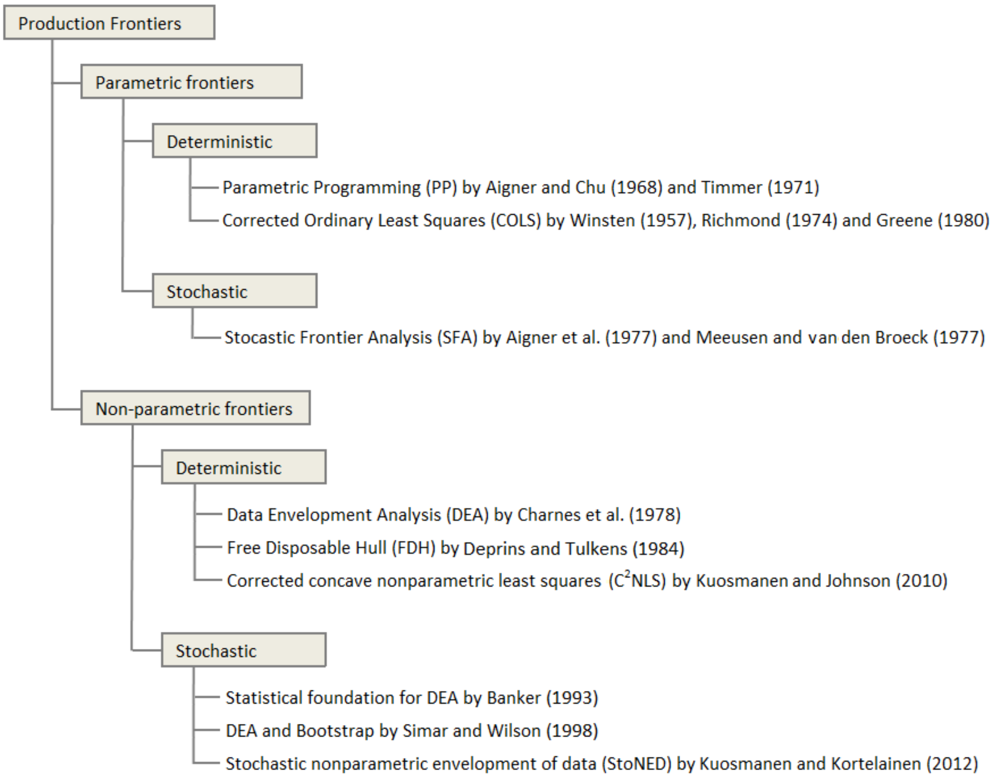


Figure 2.6: Classification of production frontiers

The stochastic approach allows for random noise and measurement error in the data. Thus, the DMUs' deviations from the estimated frontier are explained both by the DMUs' inefficiency and the presence of noise or measurement error in the data. In these cases, the estimation of the production frontier involves the use of statistical techniques.

The deterministic approach assumes that there is no random noise affecting the construction of the frontier. Thus, the DMUs' deviations from the estimated frontier are exclusively explained by inefficiency. In these cases, the estimation of the production frontier involves the use of mathematical programming techniques. As noted by Fried et al. (2008), an advantage of this approach is to avoid confounding the effects of misspecification of the functional form (as it can occur in the parametric approach), with those of inefficiency.

The remainder of the thesis will focus on the nonparametric and deterministic DEA technique, which is described in detail in the next section.

2.3 The Data Envelopment Analysis technique

The Data Envelopment Analysis (DEA) technique measures the efficiency of an homogeneous set of DMUs in their use of multiple inputs to produce multiple outputs. The efficiency is estimated by comparison to other observed DMUs, and thus it is a relative measure. DEA uses linear programming to construct a piecewise linear frontier that envelops the sample data. The measure of efficiency for each DMU is then estimated relative to the frontier constructed. The DEA method is based on the axiomatic approach outlined in the previous section, subject to certain assumptions about the structure of the production technology.

The idea of an efficiency evaluation based on observed data that accounts for multiple inputs and outputs was introduced in Farrell (1957). However, it remained unoperationalised until the paper of Charnes et al. (1978), where the term *Data Envelopment Analysis* was coined. Several DEA models were proposed in the literature after the seminal paper of Charnes et al. (1978). Some of the basic DEA models will be presented in the next sections. The mathematical representation will be introduced using the most intuitive formulation, corresponding to the ratio model.

2.3.1 DEA models

2.3.1.1 The ratio model

Consider a performance assessment involving n DMUs, $j = 1, ..., n$, each consuming m inputs x_{ij}, $i = 1, ..., m$, to produce s outputs y_{rj}, $r = 1, ..., s$. For the DMU$_{j_0}$ under assessment, it is possible to obtain a measure of relative efficiency comparing its ratio of all outputs to all inputs with similar ratios corresponding to peer DMUs (or DMUs used as comparators for the estimation of the efficiency score). The multiple inputs and outputs are reduced to a single input value and a single output value by assigning weights to each input and output. The weights are specified using an optimization procedure that aims to show the efficiency of DMU$_{j_0}$ in the best possible light.

The relative efficiency of the DMU$_{j_0}$ under assessment is obtained from the ratio model shown in (2.10).

$$
\begin{aligned}
&\text{Max } e_{j_0} = \frac{\sum_{r=1}^{s} u_r \ y_{rj_0}}{\sum_{i=1}^{m} v_i \ x_{ij_0}} && \qquad (2.10)\\
&s.t. \ \frac{\sum_{r=1}^{s} u_r \ y_{rj}}{\sum_{i=1}^{m} v_i \ x_{ij}} \leq 1 && j = 1, \ldots, n\\
&u_r \geq \epsilon && r = 1, \ldots, s\\
&v_i \geq \epsilon && i = 1, \ldots, m
\end{aligned}
$$

The variables v_i and u_r are the weights attached to the inputs and outputs, respectively. This model searches for the optimal input and output weights that maximize the efficiency of DMU$_{j_0}$, subject to the constraint that the efficiency of all DMUs in the sample is less than or equal to one, when evaluated with the same set of weights.

Note that the optimal input and output weights assigned to a given DMU under assessment may be different from the set of weights used for the evaluation of other DMUs in the sample. Individual DMUs may have their own value systems, and therefore define their own set of weights, such that their efficiency is maximized in comparison with the other DMUs in the sample.

The mathematical infinitesimal ϵ ensures that the weights are strictly positive, i.e., it ensures that all inputs and outputs are taken into account in the efficiency evaluation (Ali and Seiford (1990) provides a discussion about the choice of appropriate values of ϵ).

The ratio model presented in (2.10) can be converted into a linear programming model through simple transformations, as shown in Charnes et al. (1978). The linear programming models for efficiency evaluation using DEA are presented in the next sections.

2.3.1.2 Constant returns to scale models

The linearization of the model (2.10) leads to the DEA models shown in (2.11) and (2.12), corresponding to an input and an output orientation, respectively. Both formulations assume constant returns to scale.

For the input oriented case, the conversion of the ratio model into a linear programming model is done by maximizing the numerator of the objective function in (2.10) and setting the denominator of the objective function equal to one. For the output oriented case, the linearization is obtained by minimizing the denominator of the objective function in (2.10) and setting the numerator of the objective function equal to one.

DEA input oriented model under CRS (multiplier formulation):

$$
\begin{aligned}
&\text{Max } e_{j_0} = \sum_{r=1}^{s} u_r \; y_{rj_0} && \\
&s.t. \sum_{i=1}^{m} v_i \; x_{ij_0} = 1 && \\
&\sum_{r=1}^{s} u_r \; y_{rj} - \sum_{i=1}^{m} v_i \; x_{ij} \leq 0 && j = 1, \ldots, n \\
&u_r \geq \epsilon && r = 1, \ldots, s \\
&v_i \geq \epsilon && i = 1, \ldots, m
\end{aligned}
\tag{2.11}
$$

DEA output oriented model under CRS (multiplier formulation):

$$
\begin{aligned}
&\text{Min } h_{j_0} = \sum_{i=1}^{m} v_i \; x_{ij_0} && \\
&s.t. \sum_{r=1}^{s} u_r \; y_{rj_0} = 1 && \\
&\sum_{r=1}^{s} u_r \; y_{rj} - \sum_{i=1}^{m} v_i \; x_{ij} \leq 0 && j = 1, \ldots, n \\
&u_r \geq \epsilon && r = 1, \ldots, s \\
&v_i \geq \epsilon && i = 1, \ldots, m
\end{aligned}
\tag{2.12}
$$

Models (2.11) and (2.12) involve finding values for v_i and u_r, such that the efficiency measure for the DMU$_{j_0}$ is maximised, subject to the constraint that the efficiency measure must be less than or equal to one for all DMUs in the sample. If using the optimal weights for DMU$_{j_0}$ no other DMU reaches a value of the output to input ratio higher than the value of this ratio for DMU$_{j_0}$, it is considered efficient (in the Koopmans' sense) and is assigned a score equal to one. Otherwise, DMU$_{j_0}$ is considered inefficient. The linear programming problem is solved for each DMU, in order to allow the DMU under assessment to be assigned its own set of weights.

The relative efficiency score for DMU_{j_0} is given by $e^*_{j_0}$ in formulation (2.11) and $1/h^*_{j_0}$ in formulation (2.12). This efficiency score ranges between 0 (worst) and 1 (best). Under constant returns to scale, these models provide identical efficiency scores, i.e., $e^*_{j_0} = 1/h^*_{j_0}$.[4]

These formulations also provide further managerial information regarding the inputs or outputs associated with good performance levels for each DMU. As the weights v_i and u_r associated with the inputs and outputs, respectively, are dependent on the measurement scale of each input and output, "virtual" inputs and "virtual" outputs can be used instead to determine the relative importance attached to the inputs and outputs by each DMU, as they do not depend on the measurement scale of the variables. The virtual inputs or virtual outputs are normalised weights given by the product of the input or output value of the DMU_j with the corresponding weight ($v^*_i x_{ij_0}$ and $u^*_r y_{rj_0}$, respectively). For example, in input oriented assessments using model (2.11), the sum of the virtual weights is equal to one, so they can be interpreted as the importance, evaluated in percentage, given to each input in the estimation of the efficiency score.

The DEA linear programming models shown in (2.11) and (2.12) are known as *multiplier formulations*. By duality, these models can be expressed in their equivalent *envelopment formulations*, as shown in (2.13) and (2.14), respectively.

[4]The symbol * denotes the value of a variable at the optimal solution to the model.

DEA input oriented model under CRS (envelopment formulation):

$$
\begin{aligned}
& \text{Min } e_{j_0} = \delta - \epsilon \left(\sum_{i=1}^{m} s_i + \sum_{r=1}^{s} s_r \right) && && (2.13) \\
& s.t.\ \delta\ x_{ij_0} - \sum_{j=1}^{n} \lambda_j\ x_{ij} - s_i = 0 && i = 1, \ldots, m \\
& \sum_{j=1}^{n} \lambda_j\ y_{rj} - s_r = y_{rj_0} && r = 1, \ldots, s \\
& \lambda_j \geq 0 && j = 1, \ldots, n \\
& s_i \geq 0 && i = 1, \ldots, m \\
& s_r \geq 0 && r = 1, \ldots, s
\end{aligned}
$$

DEA output oriented model under CRS (envelopment formulation):

$$
\begin{aligned}
& \text{Max } h_{j_0} = \theta + \epsilon \left(\sum_{i=1}^{m} s_i + \sum_{r=1}^{s} s_r \right) && && (2.14) \\
& s.t.\ \sum_{j=1}^{n} \lambda_j\ x_{ij} + s_i = x_{ij_0} && i = 1, \ldots, m \\
& \theta\ y_{rj_0} - \sum_{j=1}^{n} \lambda_j\ y_{rj} + s_r = 0 && r = 1, \ldots, s \\
& \lambda_j \geq 0 && j = 1, \ldots, n \\
& s_i \geq 0 && i = 1, \ldots, m \\
& s_r \geq 0 && r = 1, \ldots, s
\end{aligned}
$$

Considering the input-oriented model shown in (2.13), the radial efficiency of DMU_{j_0} corresponds to the minimal factor (δ^*), by which its input levels can be decreased equiproportionally within the PPS, whilst the outputs are held constant. A DMU is considered radially efficient if δ^* is equal to one. For the output oriented model shown in (2.14), the radial efficiency of DMU_{j_0}

is the inverse of the maximum factor $(1/\theta^*)$ by which its outputs levels can be increased equiproportionally within the PPS, whilst the inputs are held constant. A DMU is considered radially efficient if $1/\theta^*$ is equal to one.[5]

For a DMU$_{j_0}$ to be considered truly efficient, in Koopmans' sense, it must be radially efficient, and, in addition, the slack variables s_i^* and s_r^* must be equal to zero. The slack variables indicate the extent to which each input or output can be improved beyond the amount indicated by the factors δ or θ. In order to illustrate the concept of slack, useful to understand the efficiency in Koopmans' sense, Figures 2.7 and 2.8 show the slacks in the technical efficiency measure for an input oriented and an output oriented assessment, respectively. For the seven DMUs illustrated in Figure 2.7, DMU A has a positive slack in the input 2 (S_{2A}). Conversely, DMU E and the projection of DMU G on the frontier (G*) have positive slacks in input 1 (S_{1E} and S_{1G}, respectively). To become efficient in Koopmans' sense, these DMUs should move along the segment parallel to the axes to reach the efficient frontier at points B and D, as explained in section 2.2.3.

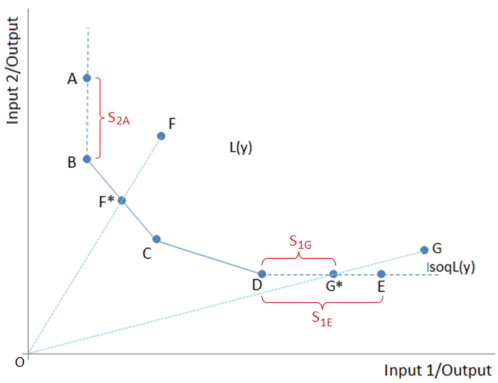


Figure 2.7: Slacks in the technical efficiency measure for an input oriented assessment

[5]Note that in the input oriented case, the radial efficiency score, given by δ^*, coincides with the Debreu-Farrell radial efficiency measure, shown in (2.6). However, in output oriented case, the radial efficiency score, given by $1/\theta^*$, corresponds to the inverse of the Debreu-Farrell radial efficiency measure, shown in (2.7).

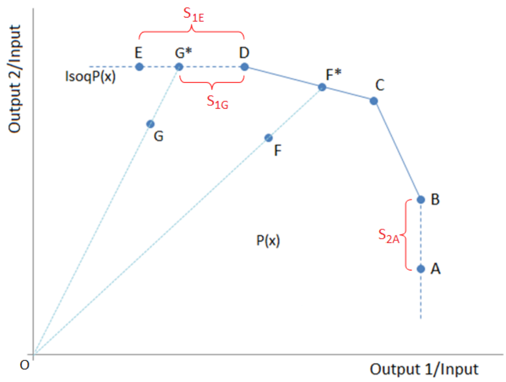


Figure 2.8: Slacks in the technical efficiency measure for an output oriented assessment

In addition to the efficiency measure, the DEA envelopment model is also able to provide other managerial information. As by-products of the efficiency assessment, it is possible to identify the peers for each inefficient DMU and the targets that the inefficient DMUs should aim to achieve in order to become efficient. For each DMU_{j_0}, the optimal solution of models (2.13) and (2.14) looks for a comparator, i.e. a composite DMU corresponding to a linear combination of efficient DMUs. This reference DMU $(\sum_{j=1}^{n} \lambda_j^* \; x_{ij}, \sum_{j=1}^{n} \lambda_j^* \; y_{rj})$ uses the same or lower levels of input and produces equal or higher levels of output than DMU_{j_0}. When $\lambda_j^* > 0$, the corresponding DMU_j is a peer to DMU_{j_0}.

The targets are given by the expressions shown in (2.15) and (2.16), corresponding to the input and output oriented models, respectively.

$$\begin{cases} x_{ij_0}^{IT} = \delta^* x_{ij_0} + s_i^* = \sum_{j=1}^{n} \lambda_j^* \; x_{ij} & i = 1, ..., m \\ y_{rj_0}^{IT} = y_{rj_0} + s_r^* = \sum_{j=1}^{n} \lambda_j^* \; y_{rj} & r = 1, ..., s \end{cases} \tag{2.15}$$

$$
\begin{cases}
x_{ij_0}^{OT} = x_{ij_0} + s_i^* = \sum_{j=1}^{n} \lambda_j^* \, x_{ij} & i = 1, ..., m \\
y_{rj_0}^{OT} = \theta^* y_{rj_0} + s_r^* = \sum_{j=1}^{n} \lambda_j^* \, y_{rj} & r = 1, ..., s
\end{cases}
\tag{2.16}
$$

The models for efficiency measurement described in this section assumed constant returns to scale. The next section will introduce models for efficiency measurement under variable returns to scale.

2.3.1.3 Variable returns to scale models

As mentioned in section 2.2.2, the returns to scale is a characteristic of the boundary of the technology of production. It measures the responsiveness of the output to equal proportional changes to the input.

The Constant Returns to Scale (CRS) assumption is appropriate when maximal productivity is attainable for all scale size ranges. However, in some cases, maximum productivity is limited to a specific scale size range and, as a consequence, some DMUs may be either too small or too large to achieve the maximum productivity. In theses cases, the appropriate assumption is the existence of variable returns to scale.

The Variable Returns to Scale (VRS) specification allows calculating technical efficiency measures accounting for scale effects. It ensures that inefficient DMUs are only compared with DMUs of similar size, such that the productivity levels observed in the peers are achievable by the DMU under assessment.

Figure 2.9 illustrates the frontiers assuming constant returns to scale and variable returns to scale for the single-output, single-input case. The CRS frontier is given by the ray starting in the origin and passing through DMU

B. Note that DMU B corresponds to the maximum productivity level of the sample. Assuming CRS, a change in the input level results in a equally proportionate change in the output level, and so, in this case, the frontier is spanned by the ray $\overrightarrow{OB}$.

By relaxing the ray unboundedness assumption used for the construction of the CRS frontier, the frontier of the PPS is defined based on the observed performance of the DMUs given their scale of operation. In the example shown in Figure 2.9, the efficient frontier is redefined as the segments between A, B and C, corresponding to a VRS frontier. The segment linking A to B exhibits increasing returns to scale (IRS), as a change to the input level causes a greater than proportionate change to the output, and the segment linking B to C exhibits decreasing returns to scale (DRS), as a change to the input causes a less than proportionate change to the output. In this case, point B, which is part both of the CRS and VRS frontiers, is the only point of the VRS frontier that is said to exhibit constant returns to scale.

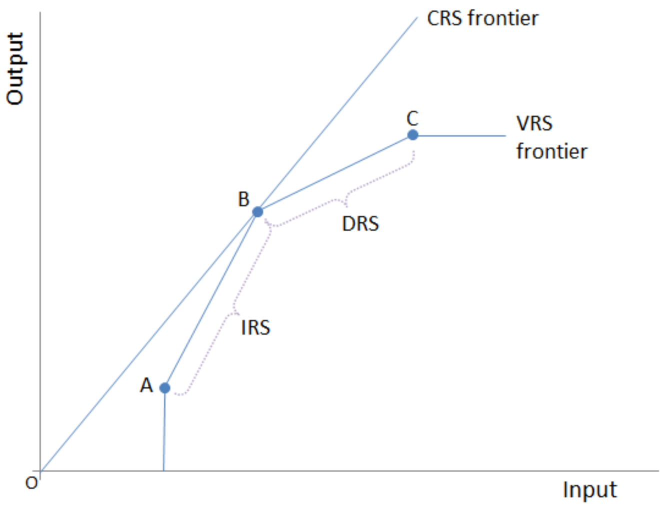


Figure 2.9: Returns to scale

Although the returns to scale concept was illustrated for the single-output, single-input case, it can be generalised to the case of multiple inputs and multiple outputs, see Banker et al. (1984).

In order to enable the estimation of efficiency assuming VRS, Banker et al. (1984) proposed a modification to the original DEA model of Charnes et al. (1978). The VRS models with input and output orientation are presented in (2.17) and (2.18), respectively.

DEA input oriented model under VRS (multiplier formulation):

$$
\begin{aligned}
&\text{Max } \hat{e}_{j_0} = \sum_{r=1}^{s} u_r \; y_{rj_0} + \omega \\
&s.t. \sum_{i=1}^{m} v_i \; x_{ij_0} = 1 \\
&\sum_{r=1}^{s} u_r \; y_{rj} - \sum_{i=1}^{m} v_i \; x_{ij} + \omega \leq 0 \qquad j = 1, \ldots, n \\
&u_r \geq \epsilon \qquad\qquad\qquad\qquad\qquad\quad r = 1, \ldots, s \\
&v_i \geq \epsilon \qquad\qquad\qquad\qquad\qquad\quad i = 1, \ldots, m \\
&\omega \text{ is free}
\end{aligned}
\tag{2.17}
$$

The variable ω can be used to impose different types of returns to scale. For non-increasing returns to scale the restriction $\omega \leq 0$ should be imposed. Conversely, for non-decreasing returns to scale the restriction $\omega \geq 0$ should be imposed.

DEA output oriented model under VRS (multiplier formulation):

$$
\begin{aligned}
&\text{Min } \hat{h}_{j0} = \sum_{i=1}^{m} v_i \; x_{ij0} + \varpi && && (2.18)\\
&s.t. \sum_{r=1}^{s} u_r \; y_{rj0} = 1 && &&\\
&-\sum_{r=1}^{s} u_r \; y_{rj} + \sum_{i=1}^{m} v_i \; x_{ij} + \varpi \geq 0 && j = 1, \ldots, n &&\\
&u_r \geq \epsilon && r = 1, \ldots, s &&\\
&v_i \geq \epsilon && i = 1, \ldots, m &&\\
&\varpi \text{ is free} && &&
\end{aligned}
$$

In model (2.18), the different types of returns to scale are imposed using the variable ϖ. The restriction $\varpi \geq 0$ is used for non-increasing returns to scale, and $\varpi \leq 0$ for non-decreasing returns to scale. [6]

Under VRS, the efficiency scores may be different depending on the model orientation considered. Thus, although the location of the frontier and the subset of efficient DMUs is the same for both model orientations, for inefficient DMUs, the scores of the input and output oriented models may be different (i.e., $e^*_{j0} \neq 1/h^*_{j0}$).

The dual models corresponding to the DEA weights formulations shown in (2.17) and (2.18) are presented in (2.19) and (2.20), respectively.

[6] Note that the value of the ω and ϖ in formulations (2.17) and (2.18) represent the intercept corresponding to the facet of the frontier against which DMU_{j0} is evaluated.

DEA input oriented model under VRS (envelopment formulation):

$$
\begin{aligned}
&\text{Min } \hat{e}_{j_0} = \hat{\delta} - \epsilon \left(\sum_{i=1}^{m} s_i + \sum_{r=1}^{s} s_r \right) && && (2.19)\\
&s.t. \ \hat{\delta}\ x_{ij_0} - \sum_{j=1}^{n} \lambda_j \ x_{ij} - s_i = 0 && i = 1, \dots, m \\
&\sum_{j=1}^{n} \lambda_j \ y_{rj} - s_r = y_{rj_0} && r = 1, \dots, s \\
&\sum_{j=1}^{n} \lambda_j = 1 \\
&\lambda_j \geq 0 && j = 1, \dots, n \\
&s_i \geq 0 && i = 1, \dots, m \\
&s_r \geq 0 && r = 1, \dots, s
\end{aligned}
$$

DEA output oriented model under VRS (envelopment formulation):

$$
\begin{aligned}
&\text{Max } \hat{h}_{j_0} = \hat{\theta} + \epsilon \left(\sum_{i=1}^{m} s_i + \sum_{r=1}^{s} s_r \right) && && (2.20)\\
&s.t. \ \sum_{j=1}^{n} \lambda_j \ x_{ij} + s_i = x_{ij_0} && i = 1, \dots, m \\
&\hat{\theta}\ y_{rj_0} - \sum_{j=1}^{n} \lambda_j \ y_{rj} + s_r = 0 && r = 1, \dots, s \\
&\sum_{j=1}^{n} \lambda_j = 1 \\
&\lambda_j \geq 0 && j = 1, \dots, n \\
&s_i \geq 0 && i = 1, \dots, m \\
&s_r \geq 0 && r = 1, \dots, s
\end{aligned}
$$

In both formulations (2.19) and (2.20), the non-increasing returns to scale assumption is obtained by imposing the sum of lambda variables to be smaller than or equal to one ($\sum_{j=1}^{n} \lambda_j \leq 1$), and the non-decreasing returns to scale is obtained by imposing the sum of lambda variables to be greater than or equal to one ($\sum_{j=1}^{n} \lambda_j \geq 1$). Regardless of the nature of the returns to scale assumed, the targets for the inefficient DMUs are obtained using the expressions presented in (2.15) and (2.16).

The nature of returns to scale of a DMU can be identified by comparing the efficiency measure derived from the different DEA models. If a DMU obtains different efficiency scores in the solution of the CRS and VRS models, then this particular DMU exhibits VRS. However, this does not indicate whether the DMU is operating in an area of increasing or decreasing returns to scale. This issue can be solved by running an additional DEA model imposing non-increasing returns to scale (NIRS). The nature of the scale inefficiencies of the DMU can be identified by comparing the efficiency scores derived from the NIRS technology and the VRS technology. If they are equal then decreasing returns to scale exist. If they are different then increasing returns to scale apply (Fare et al., 1985).

In assessments allowing for variable returns to scale, DMUs with extreme scale sizes (very small or very large) may be classified as efficient due to the lack of comparators with a similar scale size. In addition, the VRS frontier will always envelop the data as tightly as possible, regardless of whether it is the correct assumption for a given context of DMU's activity. This will result in an increase in the value of the efficiency estimate and less discrimination between the DMUs' efficiency scores whenever the VRS assumption is used. Given these limitations, VRS should be assumed only when there are strong evidences that scale effects actually exist.

In cases in which the nature of the production technology (CRS or VRS) is unknown *a priori*, hypothesis tests can be used to verify the existence of scale effects. The VRS model should only be used when scale effects are demonstrated. Banker (1996) was the first to use hypothesis tests in a DEA assessment to verify the nature of the returns to scales. Some years later, Simar and Wilson (2002) proposed the use of bootstrapping to test the hypothesis regarding returns to scale.

Next, the concept of scale efficiency is introduced. Scale efficiency is a measure of how much the scale of operation of a DMU impacts on its ability to achieve the maximum productivity. Comparing the distance between the CRS and VRS frontiers corresponding to the evaluation of a given DMU, it is possible to define a measure of scale efficiency. The scale efficiency of DMU_{j_0} can be calculated as shown (2.21).

$$\text{Scale efficiency of DMU}_{j_0} = \frac{\text{CRS efficiency score of DMU}_{j_0}}{\text{VRS efficiency score of DMU}_{j_0}}. \tag{2.21}$$

The concept of scale efficiency is illustrated in Figure 2.10 adopting an output orientation for the efficiency assessment. DMU D is not technically efficient as it is operating below the efficient frontier. It could become efficient, in a pure technical sense, by increasing its output level until reaching point D^*. However, at this point, DMU D would be considered scale inefficient because it is not possible to achieve the maximum productivity level (observed at DMU B). The scale efficiency measure, given by the ratio $\frac{O'D^*}{O'D^{**}}$, evaluates the distance between the CRS and VRS frontiers, and measures the amount of output loss attributable to having a scale size that prevents attaining maximum productivity. Therefore, the overall efficiency measure for DMU D, which incorporates both pure technical and scale efficiencies, is given by the product of the pure technical and scale efficiency scores, i.e., $\frac{O'D}{O'D^*} \times \frac{O'D^*}{O'D^{**}} = \frac{O'D}{O'D^{**}}$.

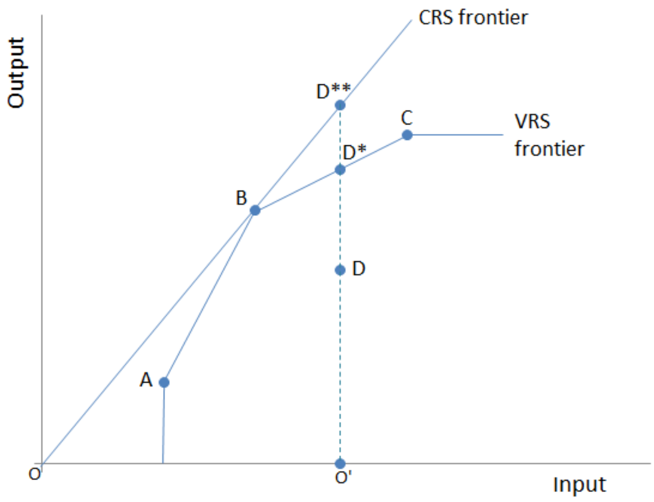


Figure 2.10: Scale efficiency measurement

2.3.2 Restricting weights in DEA models

The original DEA model, developed by Charnes et al. (1978), allows total flexibility in the selection of the weights v_i and u_r to be attached to the inputs and outputs, such that each DMU tries to maximise its efficiency score, given the inputs consumed and output levels attained. This total flexibility in the selection of weights is important for the identification of under-performing DMUs. Under these conditions, if a DMU does not achieve the maximum score, even when evaluated with a set of weights that intends to maximise its performance score, it provides irrefutable evidence that other DMUs are performing better. However, this feature can also cause problems. For example, one can argue that the weights assigned to the inputs and outputs are not realistic, and thus the robustness of the efficiency measure and its applicability in real-world context is questionable.

The first attempt to restrict the flexibility of the weights in DEA models was made in the mid 80's. The use of DEA by Thompson et al. (1986) to support the selection of potential sites for a nuclear research laboratory in Texas raised a problem of lack of discrimination between efficient DMUs. The authors improved the discrimination between the DMUs' efficiency scores by defining ranges of acceptable weights. In addition to the possibility of improving discrimination between efficient DMUs, Allen et al. (1997) pointed out a number of other reasons that can motivate the use of weight restrictions in DEA. These include the need of incorporating prior views (or value judgments) on the importance of individual inputs and outputs, relating the values of certain inputs and/or outputs, and respecting the economic notion of input or output substitution.

Since the work of Thompson et al. (1986), several types of weight restrictions have been proposed in the DEA literature. Allen et al. (1997) and Thanassoulis et al. (2004) reviewed the literature on weight restrictions and discussed the advantages and limitations of the different approaches. The weight restrictions that can be included in DEA models can be classified as *direct restrictions to weights* and *restrictions to virtual weights*. These will be used throughout the thesis, and are explained in detail in the following sections.

2.3.2.1 Direct restrictions to weights

There are three types of direct restrictions to weights, namely *Absolute weight restrictions*, *Assurance Regions type I* (ARI) and *Assurance Regions type II* (ARII).

The absolute weight restrictions were first introduced by Dyson and Thanassoulis (1988). They are included in the DEA model in the form shown in (2.22) and (2.23), corresponding to restrictions associated with input weights v_i and output weights u_r, respectively.

$$\varsigma_i \leq v_i \leq \tau_i \tag{2.22}$$

$$\rho_r \leq u_r \leq \eta_r \tag{2.23}$$

The parameters ς_i and τ_i in (2.22) and ρ_r and η_r in (2.23) represent the bounds to the values of the input weights and output weights, respectively, that are considered acceptable. This type of weight restrictions is usually introduced to prevent the inputs or outputs from being over emphasised or ignored in the assessment.

The absolute weight restrictions shown in (2.22) and (2.23) may appear to be a simple form of weight restrictions. However, a number of difficulties are associated with them. When absolute weight restrictions are imposed in DEA models, they may render infeasible solutions or lead to the underestimation of the efficiency scores (Allen et al., 1997; Podinovski and Athanassopoulos, 1998).

Another difficulty associated with absolute weight restrictions is the meaning of the bounds (ς_i, τ_i, ρ_r and η_r). In general, weights are meaningful only on a relative basis (Fried et al., 2008). Depending on the context, however, it may be possible to attribute a meaning to the bounds. For example, in situations in which the DMUs under assessment consume a single input, Dyson and Thanassoulis (1988) interpreted the weight on the r^{th} output as the amount of resource the DMU is deemed to consume per unit of output r. However, this interpretation does not readily extend to the case of multiple inputs.

Dyson et al. (2001) noted that absolute weight restrictions are not directly transferable between models. For example, the ratio model (2.10) with the absolute bounds (2.22) and (2.23) is generally not equivalent to the linear forms (2.11) and (2.12) with the same bounds, so the efficiency of the DMUs may be different among these models. In addition, as pointed out by Allen et al. (1997), switching from an input to an output orientation produces different relative efficiency scores and hence the bounds need to be set considering the model orientation used.

In order to avoid these problems and obtain a correct estimate of relative efficiency in the presence of absolute weight restrictions, Podinovski and Athanassopoulos (1998) proposed the use of a Maximin model, and developed an equivalent linear programming formulation to enable the computation of relative efficiency scores.

Assurance Regions differ from absolute weight restrictions because instead of requiring the weights to be within certain limits, they require ratios between weights to be within certain limits. The most prevalent type of direct weight restrictions used in DEA applications are assurance regions type I (ARI), proposed by Thompson et al. (1990). They usually incorporate information concerning marginal rates of substitution between inputs or outputs, as shown in formulations (2.24) and (2.25), respectively.

The parameters ϱ_i and π_i in (2.24) and χ_r and ζ_r in (2.25) correspond to the bounds that the ratio of input and output weights, respectively, can assume.

$$\varrho_i \leq \frac{v_i}{v_{i+1}} \leq \pi_i \tag{2.24}$$

$$\chi_r \leq \frac{u_r}{u_{r+1}} \leq \zeta_r \tag{2.25}$$

As pointed out by Allen et al. (1997) and Sarrico and Dyson (2004), a disadvantage of this type of weight restrictions is that they are sensitive to the units of measurement of inputs and outputs. As a result, it is often difficult to specify meaningful marginal rates of substitution between the variables. Unlike the absolute weight restrictions, when assurance regions type I are added to CRS models, the efficiency scores obtained are identical for input and output oriented models, as well as for the ratio model.

The Assurance Regions type II (ARII), introduced by Thompson et al. (1990), are used to reflect relationships between input and output weights. Thus, while the ARI are used to specify ratios either between input or output weights separately, and ARII are used to specify ratios that link input to output weights. They are expressed in the form shown in (2.26).

$$\gamma_i v_i \geq u_r \tag{2.26}$$

Similarly to the ARIs, the ARIIs are sensitive to the units of measurement of the inputs and outputs.

In addition, similarly to the absolute weight restrictions, problems related to infeasible solutions and under-estimation of the relative efficiency scores may occur when ARII are added to DEA models. This may justify the small number of empirical applications using ARII in DEA models. As noted by Khalili et al. (2010), the use of ARII is more frequent in profit ratio models, which differ from standard DEA models, as in these last models the constraints imposing the ratio of virtual outputs to virtual inputs to be smaller than or equal to one are removed for all DMUs.

In order to overcome some limitations of the ARII, Thompson and Thrall (1994) introduced a nonlinear DEA model that can retrieve the correct relative efficiency scores in the presence of ARII. However, this model was only solved for the special case of a single output and two inputs. Khalili et al. (2010) also addressed the problems of using DEA models with ARII. The authors proposed an alternative nonlinear model that can successfully measure relative efficiency in the presence of ARII. This model is inspired on the idea of measuring relative efficiency through the Maximin model, first proposed by Podinovski and Athanassopoulos (1998).

2.3.2.2 Restrictions to virtual weights

The virtual outputs (inputs) are the product of the output (input) value of the DMU$_j$ with the corresponding weight u_r (v_i). The restrictions to virtual outputs or inputs, called virtual weight restrictions, were originally proposed by Wong and Beasley (1990). Such restrictions assume the form presented in (2.27). They restrict the importance attached to the output indicator y_r, expressed in percentual terms, ranging between a lower and an upper bound (ϕ_r and ψ_r, respectively). A similar restriction can be set to virtual inputs.

$$\phi_r \leq \frac{u_r \ y_{rj}}{\sum_{r=1}^{s} u_r \ y_{rj}} \leq \psi_r \qquad j = 1, ..., n \tag{2.27}$$

A key advantage of restrictions to virtual weights is that they are independent of the units of measurement of the inputs and outputs. However, as pointed out by Thanassoulis et al. (2004), the restrictions to virtual weights are DMU-specific, so they may be computationally expensive. They may also lead to infeasible solutions when the bounds are loosely specified. As suggested by Wong and Beasley (1990), an alternative to overcome the infeasibility problems and the computational difficulties, is to apply the above restrictions only to the virtual outputs of the DMU$_{j0}$ under assessment. However, the restrictions imposed only to the DMU under assessment also have drawbacks. According to Dyson et al. (2001), if the restrictions are imposed only to the virtual outputs of the DMU under assessment, they compromise the symmetry of the model with respect to all DMUs, as each DMU is assessed based on a different feasible region. Sarrico and Dyson

(2004) added that these restrictions imposed only to the DMU under assessment might impose unreasonable restrictions on the virtual weights of the other DMUs.

2.4 Summary of the chapter

This chapter provided an overview of the theory of production, as it constitutes the foundations for the estimation of efficiency. The assumptions underlying the construction of the production possibility set were reviewed. In addition, the different measures of efficiency were described, along with the concept of returns to scale. The DEA technique was discussed in detail, as this is the main approach used throughout the thesis. Finally, the incorporation of values judgments in DEA models in the form of weight restrictions was outlined.

Recent developments corresponding to the construction of composite indicators using DEA, and the treatment of undesirable outputs in this context will be reviewed and discussed in the next chapters, where the contributions of the thesis related to these topics are presented.

Chapter 3

Modeling undesirable outputs in the construction of DEA-based composite indicators

3.1 Introduction

In standard Data Envelopment Analysis models, as presented in chapter 2, an inefficient DMU can improve its performance by increasing the levels of outputs (results obtained) or decreasing the levels of inputs (resources used). This point of view makes sense when all outputs are intended to be increased and all inputs are intended to be reduced. However, real-world applications may involve both desirable and undesirable outputs and inputs. The accommodation of undesirable factors in DEA models is not straightforward.

Although the issue of dealing with undesirable outputs and inputs in DEA models has been approached in the literature by various authors, they do not address the modeling of undesirable factors in the construction of composite indicators. A composite indicator is given by the aggregation of several individual indicators, it focuses only in the achievements (results obtained) of a set of DMUs and are intended to reflect multidimensional concepts in a single measure. In composite indicators constructed using DEA models, all individual indicators are specified as outputs and a unitary input underlying the evaluation of every DMU is considered.

The use of DEA for performance assessments focusing only on achievements, rather than the conversion of inputs to outputs, was first proposed by Cook and Kress (1990), with the purpose to construct a preference voting model (for aggregating votes in a preferential ballot). Other relevant studies that support the empirical use of DEA models only with outputs (or productivity indicators that aggregate output and input information, such as revenue per employee and GDP per capita) can be found in different fields, such as macroeconomic performance assessment (Lovell et al., 1995), human development (Mahlberg and Obersteiner, 2001; Despotis, 2004, 2005), technology achievement (Cherchye et al., 2008) and evaluation of urban quality of life (Morais and Camanho, 2011). In these studies, all variables were specified as outputs and an identical input level, which for simplicity was assumed to be equal to one, was specified for all DMUs.

In this chapter, we approach two main issues: the construction of CIs that include both desirable and undesirable factors with aggregation procedures based on DEA, and the use of weight restrictions in this context.

Two aggregation procedures are considered to construct the CI in the presence of undesirable outputs. First, the CI is derived based on a traditional DEA model. In this case the undesirable outputs require a prior transformation in the measurement scale to be accommodated in the CI model. Next, we propose the use of a CI model specified using a directional distance function. This CI model is able to accommodate the undesirable outputs in their original form. The features, weaknesses and advantages of both approaches are discussed and illustrated using a small example.

Concerning the use of weight restrictions in the context of the estimation of CIs, it is explored the implementation of the two most popular types of weight restrictions: virtual weight restrictions and the assurance regions type I (ARIs). We propose an enhanced formulation of weight restrictions to incorporate the relative importance of individual indicators expressed in percentage, using ARI. We discuss and illustrate the features of each type of weight restriction using a small example.

Therefore, from a methodological perspective, the two major contributions of this chapter consist on the construction of DEA-based CIs that can accommodate both desirable and undesirable outputs and provide peers and targets as by-products of the assessment, and the specification of a novel type of weight restriction, using ARIs, to incorporate the relative importance of indicators, expressed as a percentage.

The remainder of this chapter proceeds as follows. Section 3.2 provides a literature review of studies that have approached the issue of dealing with undesirable outputs in DEA models. Section 3.3 presents the DEA formulations that can be used for efficiency assessments in the presence of undesirable outputs, and adapts the models to evaluations with composite indicators. Section 3.4 approaches the construction of composite indicators incorporating decision-maker preferences, using weight restrictions of different types. Section 3.5 illustrates the specificities of the models, and discusses their strengths and weaknesses using a small numerical example. Finally, section 3.6 presents the conclusions.

3.2 Review of the literature on undesirable outputs in DEA

Although the seminal paper of Koopmans (1951) already mentioned that the production process may also generate undesirable outputs, such as air pollutants or waste, it was only in the 80s that the issue of efficiency measurement in the presence of undesirable outputs was first addressed. One of the earliest studies addressing the incorporation of undesirable outputs in the assessment of production efficiency was developed by Pittman (1983). This study extended the multilateral productivity indicator proposed by Caves et al. (1982) to include measures of both desirable and undesirable outputs. The multilateral productivity indicator developed by Caves et al. (1982) required the specification of the price data, but this information is often unavailable for undesirable factors. Therefore, Pittman (1983) proposed an extension of this indicator, which assigned a value to the undesirable outputs based on estimates of shadow prices instead of market prices. Some years later, Fare et al. (1993) proposed an alternative method to estimate shadow prices based on the distance function defined by Shephard (1970). The specification of the shadow prices of undesirable outputs using a linear programming model allowed enhancing the approach proposed by Pittman (1983).

Fare et al. (1989) also proposed a modification of Farrell (1957) approach to efficiency measurement to allow an asymmetric treatment of desirable and undesirable outputs. While the multilateral productivity indicator requires the specification of price information for the undesirable outputs, the non-parametric approach of Fare et al. (1989) only requires data on quantities of the undesirable outputs. The authors proposed a hyperbolic model to efficiency measurement to allow considering different assumptions on the disposability of undesirable outputs. The new constraints state that the desirable outputs are strongly disposable (i.e. they can be reduced without cost), while the undesirable outputs are weakly disposable (i.e. they can only be reduced in conjunction with a reduction in the other outputs or an increase in the use of inputs).

Some years later, Chung et al. (1997) introduced a different approach to deal with undesirable outputs in the efficiency and productivity measurement literature. The authors extended the Chambers et al. (1996) directional distance function to allow expanding the desirable outputs while simultaneously contracting the undesirable ones. The outputs are expanded or contracted along a path that is defined according to a directional vector. The directional distance function has been widely used in the context of environmental performance assessments, in which the production of waste (an undesirable output) is often present.

The approaches mentioned above are known as direct approaches to treat undesirable outputs. These approaches allow treating the outputs in their original form, that is, without requiring any modification to the measurement scale. On the other hand, there are indirect approaches that transform the values of the undesirable outputs to allow treating them as normal outputs in traditional DEA models.

Scheel (2001), Dyson et al. (2001) and Seiford and Zhu (2002) discussed the different approaches to handle undesirable outputs in DEA models using indirect approaches. One option is to move the variables from the output to the input side. Scheel (2001) pointed out that this approach results in the same technology set as incorporating the undesirable outputs as normal outputs, in the form of their additive inverses ($-y_{und}$). The incorporation of the undesirable outputs in the form of their additive inverses was first suggested by Koopmans (1951). Regarding this option, Seiford and Zhu (2002) pointed out that to treat undesirable outputs as inputs would not reflect the real production process, as the input-output structure that defines the production process would be lost. Another possibility is to consider the undesirable outputs in the form of their multiplicative inverses ($1/y_{und}$), as proposed by Golany and Roll (1989). Regarding this option, Dyson et al. (2001) pointed out that this transformation would destroy the ratio or interval scale of the data. The third option is to add to the additive inverses of the undesirable outputs a sufficient large positive number ($-y_{und} + c$), as first suggested by Ali and Seiford (1990). This transformation is the most frequently used in the literature to deal with undesirable outputs using a traditional DEA formulation (Cook and Green, 2005; Oggioni et al., 2011). It has the advantage of enabling a simple interpretation of results, but it is sensitive to the choice of the constant c, as will be discussed in the next section.

In addition to the above mentioned approaches, in Cherchye et al. (2011) the transformation in the measurement scale of the undesirable outputs was performed based on a normalization procedure, which was applied both to desirable and undesirable outputs. This procedure results in indicators varying between 0 and 1. As data normalization leads to a loss of information, this approach is rarely used in DEA studies. It does not take advantage of the ability of DEA to deal with data measured on different scales.

3.3 Undesirable outputs in the construction of DEA-based composite indicators

In this section we discuss the main models available to treat undesirable outputs in DEA efficiency assessments. This is followed by the presentation of CI models that can be obtained based on these DEA formulations. As mentioned in the literature review, the models can follow a direct or an indirect approach to handle the undesirable outputs.

3.3.1 Indirect approach

The DEA output oriented model by Charnes et al. (1978), presented in section 2.3.1.2, can be used for assessments involving undesirable outputs with a transformation in the measurement scale of the undesirable outputs as proposed by Seiford and Zhu (2002). The resulting model, with a constant returns to scale formulation in shown in (3.1).

$$
\begin{aligned}
&\text{Min } h_{j_0} = \sum_{i=1}^{m} v_i \; x_{ij_0} && && (3.1)\\
&s.t. \sum_{r=1}^{s} u_r \; y_{rj_0} + \sum_{k=1}^{l} p_k \; (M_k - b_{kj_0}) = 1 && &&\\
&\sum_{r=1}^{s} u_r \; y_{rj} + \sum_{k=1}^{l} p_k \; (M_k - b_{kj}) - \sum_{i=1}^{m} v_i \; x_{ij} \leq 0 && j = 1, \ldots, n &&\\
&u_r \geq 0 && r = 1, \ldots, s &&\\
&p_k \geq 0 && k = 1, \ldots, l &&\\
&v_i \geq 0 && i = 1, \ldots, m &&
\end{aligned}
$$

In formulation (3.1), the efficiency score for DMU j_0 is given by $1/h_{j_0}$, and it ranges between zero (worst) and one (best). x_{ij} $(i = 1, ..., m)$, y_{rj} $(r = 1, ..., s)$ and b_{kj} $(k = 1, ..., l)$ correspond to the value of the input i, desirable output r and undesirable output k, respectively, for DMU j $(j = 1, ..., n)$. v_i, u_r and p_k are the weights attached to the inputs, desirable outputs and undesirable outputs, respectively, in the performance assessment. M_k is a large positive number greater than or equal to the maximum value of the undesirable output k observed in all DMUs. In the transformation proposed by Seiford and Zhu (2002), all of the translated outputs are positive, so the constant M_k must be a value larger than the maximum observed for each output indicator. As the model is sensitive to the choice of the M_k

value, a sensitivity analysis of the results for different values of M_k should be undertaken.

In a CI we have only outputs to be aggregated, so we can assume that all DMUs are similar in terms of inputs. Thus, following Koopmans (1951), Lovell (1995) and Lovell et al. (1995), we can have a unitary input underlying the evaluation of every DMU, interpreted as a "helmsman" attempting to steer the DMUs towards the maximization of outputs. By considering a unitary input level for all DMUs in model (3.1) we obtain the CI model presented in (3.2).

$$\begin{aligned}
&\text{Min } h_{j_0} = v && \qquad (3.2)\\
&s.t. \sum_{r=1}^{s} u_r \; y_{rj_0} + \sum_{k=1}^{l} p_k \; (M_k - b_{kj_0}) = 1 \\
&\sum_{r=1}^{s} u_r \; y_{rj} + \sum_{k=1}^{l} p_k \; (M_k - b_{kj}) - v \leq 0 \quad && j = 1, \ldots, n \\
&u_r \geq 0 && r = 1, \ldots, s \\
&p_k \geq 0 && k = 1, \ldots, l \\
&v \geq 0
\end{aligned}$$

In formulation (3.2), the reciprocal of the value of the objective function represents an efficiency measure that ranges between zero and one, where one correspond to the best level of performance observed in the sample. This is also the value of the composite indicator associated with this model that will be used throughout the thesis.

The use of DEA to construct CIs was popularized by Cherchye et al. (2007). This approach is known as "benefit of the doubt" construction of composite indicators. The CI proposed by Cherchye et al. (2007) is equivalent to the input oriented model proposed by Charnes et al. (1978), assuming constant returns to scale and a unitary input level for all DMUs. A CI can also be obtained from the output oriented model, and this was the approach followed in this chapter, leading to the formulation (3.2). As both formulations assume constant return to scale, the performance scores obtained with different orientations are the same. The advantage of using the output oriented formulation, as done in this chapter, is that it leads to a more direct estimation of targets using the dual model, and facilitates the incorporation of weight restrictions.

For benchmarking purposes, the identification of the peers and targets for the inefficient DMUs can be done through the envelopment formulation of model (3.2), shown in (3.3). The objective function value at the optimal solution of the model (3.3) corresponds to the factor θ by which all outputs of the DMU under assessment can be proportionally improved to reach the target output values. As occurred in the primal formulation shown in (3.2), the performance score, or composite indicator, of DMU j_0 under assessment is the reciprocal of the objective function value of model (3.3). Therefore, the DMUs with the best performance are those for which there is no evidence that it is possible to expand their outputs, such that the value of θ^* is equal to 1.

$$
\begin{aligned}
&\text{Max } \theta && && (3.3)\\
&s.t.\;\; \theta\, y_{rj_0} - \sum_{j=1}^{n} \lambda_j\, y_{rj} \leq 0 && r = 1, \ldots, s\\
&\theta\,(M_k - b_{kj_0}) - \sum_{j=1}^{n} \lambda_j\,(M_k - b_{kj}) \leq 0 && k = 1, \ldots, l\\
&\sum_{j=1}^{n} \lambda_j \leq 1 && j = 1, \ldots, n\\
&\lambda_j \geq 0 && j = 1, \ldots, n
\end{aligned}
$$

The peers for the DMU j_0 under assessment are the DMUs with values of λ_j^* greater than zero at the optimal solution of model (3.3). The targets that a DMU j_0, with a composite indicator score smaller than one, should reach to improve performance are given by the following expressions:

$$\begin{cases} {y_{rj_0}}^{target} = \sum_{j=1}^{n} \lambda_j^* \; y_{rj} & r = 1, ..., s \\ {b_{kj_0}}^{target} = M_k - \sum_{j=1}^{n} \lambda_j^* \; (M_k - b_{kj}) & k = 1, ..., l \end{cases} \tag{3.4}$$

3.3.2 Direct approach

The efficiency evaluation using the directional distance function (DDF), developed by Chambers et al. (1996), allows to simultaneously expand outputs and contract inputs according to a directional vector. Chung et al. (1997) extended this approach to allow including undesirable outputs in the efficiency evaluation. The Chung et al. (1997) model also assumes weak disposability of undesirable outputs, as proposed in Fare et al. (1989), but preserves the linearity of the DEA model. The constant returns to scale model of Chung et al. (1997) is specified as shown in (3.5).

$$\begin{aligned} &\text{Max } \beta && \qquad (3.5) \\ &s.t. \sum_{j=1}^{n} y_{rj} \; \lambda_j \geq y_{rj_0} + \beta \; g_{y_{rj_0}} && r = 1, \dots, s \\ &\sum_{j=1}^{n} b_{kj} \; \lambda_j = b_{kj_0} - \beta \; g_{b_{kj_0}} && k = 1, \dots, l \\ &\sum_{j=1}^{n} x_{ij} \; \lambda_j \leq x_{rj_0} - \beta \; g_{x_{ij_0}} && i = 1, \dots, m \\ &\lambda_j \geq 0 && j = 1, \dots, n \end{aligned}$$

In formulation (3.5), x_{ij} $(i = 1, ..., m)$ are the inputs used by the DMU j $(j = 1, ..., n)$ to produce y_{rj} $(r = 1, ..., s)$ desirable outputs and b_{kj} $(k = 1, ..., l)$ undesirable outputs. The λ_j are the intensity variables. The components of vector $g = (g_y, -g_b, -g_x)$ indicate the direction of change for the outputs and inputs. Positive values for the components are associated with expansion of desirable outputs and negative values are associated with contraction of inputs and undesirable outputs. The factor β indicates the extent of DMU's inefficiency. It corresponds to the maximal feasible expansion of desirable outputs and contraction of inputs and undesirable outputs that can be achieved simultaneously.

While inputs and desirable outputs are assumed to be strongly disposable, the undesirable outputs are assumed to be weakly disposable, as it is shown by the equality in the constraint associated with the undesirable outputs. When imposing weak disposability of undesirable outputs we are assuming that they are by-products of the desirable outputs and cannot be reduced without cost, which implies that abatement in an undesirable output is possible if accompanied by a reduction in a desirable output or an increase in an input. The decision on whether to assume strong disposability or weak disposability for the variables of a DEA model depends on the nature of the application under analysis (Liu et al., 2009).

Using a unitary level of input and setting the directional vector as $g = (g_y, -g_b, 0)$, a vector that allows to, simultaneously, expand the desirable outputs and contract undesirable ones by keeping inputs fixed, the input restriction becomes $\sum_{j=1}^{n} \lambda_j \leq 1$. So, in the context of performance evaluations based on CIs, model (3.5) reduces to the formulation shown in (3.6). In the remainder of this chapter, we refer to this formulation as Directional CI model.

$$
\begin{aligned}
& \text{Max } \beta && && (3.6) \\
& s.t. \sum_{j=1}^{n} y_{rj}\ \lambda_j \geq y_{rj_0} + \beta\ g_{y_{rj_0}} && r = 1, \ldots, s \\
& \sum_{j=1}^{n} b_{kj}\ \lambda_j = b_{kj_0} - \beta\ g_{b_{kj_0}} && k = 1, \ldots, l \\
& \sum_{j=1}^{n} \lambda_j \leq 1 \\
& \lambda_j \geq 0 && j = 1, \ldots, n
\end{aligned}
$$

Chung et al. (1997) explained that when the directional vector is specified as the current value of the outputs for the DMU under assessment, i.e., $g = (g_y, -g_b) = (y_{rj_0}, -b_{tj_0})$, the directional distance function is comparable to the Shephard's output distance function and thus the output oriented efficiency measure is given by $1/(1 + \beta^*)$. This value ranges between zero and one, where one corresponds to the best level of performance observed in the sample. This efficiency score will also be used throughout this thesis as the value of the CI obtained using model (3.6).

As by-products of the assessment using the Directional CI model, it is possible to identify the peers and targets for the inefficient DMUs. The peers are the DMUs with λ_j^* greater than zero at the optimal solution, and the targets are given by the expressions shown in (3.7).

$$
\begin{cases}
{y_{rj_0}}^{target} = \sum_{j=1}^{n} \lambda_j^*\ y_{rj} & r = 1, ..., s \\
{b_{kj_0}}^{target} = \sum_{j=1}^{n} \lambda_j^*\ b_{kj} & k = 1, ..., l
\end{cases}
\qquad (3.7)
$$

The dual of the model (3.6), is shown in the multiplier formulation (3.8). This formulation is preferable in the case of assessments involving the incorporation of weight restrictions.

$$\begin{aligned}
&\min \quad -\sum_{r=1}^{s} y_{rj_0}\ u_r + \sum_{k=1}^{l} b_{kj_0}\ p_k + v && && (3.8)\\
&s.t. \sum_{r=1}^{s} g_{y_{rj_0}}\ u_r + \sum_{k=1}^{l} g_{b_{kj_0}}\ p_k = 1 && \\
&-\sum_{r=1}^{s} y_{rj}\ u_r + \sum_{k=1}^{l} b_{kj}\ p_k + v \geq 0 && j = 1, \ldots, n \\
&u_r \geq 0 && r = 1, \ldots, s \\
&p_k \ is\ free && k = 1, \ldots, l \\
&v \geq 0 &&
\end{aligned}$$

The objective function value at the optimal solution of the model (3.8) corresponds to the maximal feasible expansion of desirable outputs and contraction of undesirable outputs that can be achieved simultaneously.

3.4 Incorporating value judgments in composite indicators

Although DEA allows the specification of the indicator weights recurring to optimization, in some cases it may be important to incorporate in the model expert opinion about the weight that each individual indicator should have in the assessment. This can be done by imposing restrictions to weights in the DEA model. As noted by Cherchye et al. (2011), the ability to add extra information related to the importance of the individual indicators enables enhancing credibility and acceptance of CIs in practical applications. This section discusses the implementation of restrictions to weights in the context of assessments involving composite indicators.

As reviewed in the previous chapter, the weight restrictions can be classified as direct restriction to weights and virtual weight restrictions. While the direct restrictions to weights are more often used in DEA assessments that involve an input-output framework, the restrictions to virtual weights are more prevalent in assessments with composite indicators. Cherchye et al. (2007) presented and discussed different ways of implementing the virtual weight restrictions in this context.

3.4.1 Restrictions to virtual weights in CIs

As explained in section 2.3.2.2, the restrictions to virtual weights were first proposed by Wong and Beasley (1990), and they assume the form presented in (3.9). They restrict the importance attached to the output indicator y_r, expressed in percentual terms, ranging between a lower and an upper bound (ϕ_r and ψ_r, respectively).

$$\phi_r \leq \frac{u_r\ y_{rj}}{\sum_{r=1}^{s} u_r\ y_{rj}} \leq \psi_r \quad j = 1, \ldots, n \tag{3.9}$$

Cherchye et al. (2007) pointed out that CIs are often composed by individual indicators that can be classified in mutually exclusive categories C_z, $z = 1, ..., q$. In this case, one may want to impose restrictions on a category of indicators rather than on individual output indicators, specially when it is difficult to define weights for individual indicators. The restrictions to categories of indicators are natural extensions of the restrictions presented in (3.9). Instead of restricting the relative importance allocated to the output indicator y_r, they restrict the relative importance allocated to the set of output indicators from category C_z, as shown in (3.10).

$$\phi_z \leq \frac{\sum_{r \in C_z} u_r\ y_{rj}}{\sum_{r=1}^{s} u_r\ y_{rj}} \leq \psi_z \quad j = 1, \ldots, n \tag{3.10}$$

An important advantage of the restrictions to virtual weights is that they are independent of the units of measurement of the inputs and outputs. However, as they are DMU-specific, they may be computationally expensive and lead to infeasible solutions when the bounds are loosely specified. In order to overcome this problems, Wong and Beasley (1990) suggested to apply the above restrictions only to the virtual outputs of the DMU under assessment (j_0), as shown in (3.11). This procedure has been often used in the literature (e.g. Morais and Camanho (2011); Lins et al. (2012); Rogge (2012)). However, the restrictions imposed only to the DMU under assessment also have drawbacks related to the asymmetry of the model with respect to all DMUs and the possibility to obtain unreasonable restrictions on the virtual weights of the other DMUs, as discussed in more detail in section 2.3.2.2.

$$\phi_r \leq \frac{u_r \; y_{rj_0}}{\sum_{r=1}^{s} u_r \; y_{rj_0}} \leq \psi_r \tag{3.11}$$

When the DEA model is output oriented, if the restrictions to the virtual weights are only imposed to the DMU under assessment, the denominator of expression (3.11) corresponds to the normalization constraint of the DEA model, which is always equal to 1. Thus, expression (3.11) can be simplified, as shown in (3.12).

$$\phi_r \leq u_r \; y_{rj_0} \leq \psi_r \tag{3.12}$$

The virtual weight restrictions can also be included in the CI models (3.2) and (3.8). For the CI model (3.2) the virtual weight restrictions that can be imposed to the desirable outputs (y_r) and undesirable outputs (p_k) are shown in (3.13).

$$\begin{cases} \phi_r \leq u_r \; y_{rj_0} \leq \psi_r \\ \phi_k \leq p_k \; (M_k - b_{kj_0}) \leq \psi_k \end{cases} \tag{3.13}$$

Similarly, for the CI model (3.8) the specification of the output virtual weight restrictions are shown in (3.14).

$$\begin{cases} \phi_r \leq u_r \ y_{rj_0} \leq \psi_r \\ \phi_k \leq p_k \ b_{kj_0} \leq \psi_k \end{cases} \tag{3.14}$$

3.4.2 ARI restrictions in CIs

The most prevalent type of direct weight restrictions used in DEA applications are assurance regions type I (ARI), proposed by Thompson et al. (1990). They usually incorporate information concerning marginal rates of substitution between the outputs (or between the inputs), as explained in section 2.3.2.1.

It is worth highlighting that this type of weight restrictions is sensitive to the units of measurement of the inputs and outputs (Allen et al., 1997). As a result, it is often difficult to specify meaningful marginal rates of substitution between the variables in empirical applications.

In this chapter we propose an enhanced formulation of ARI weight restrictions that enables expressing the relative importance of the output indicators in percentual terms, instead of specifying marginal rates of substitution. This requires the use of an "artificial" DMU representing the average values of the outputs in the sample analyzed. This type of formulation for the weight restrictions recurring to the use of an "artificial" DMU, equal to the sample average, was originally proposed by Wong and Beasley (1990) as a complement of DMU-specific virtual weight restrictions.

If instead of restricting the virtual outputs of a DMU$_j$, as shown in (3.9), the restrictions are imposed to the average DMU ($\bar{y}_r$), as shown in (3.15), all DMUs are assessed with identical restrictions[1]. Thus, these weight re-

[1] Note that using the restrictions shown in (3.11) each DMU$_{j_0}$ is assessed with its own weight restrictions, whose bounds depend on the value of the output y_r observed for each DMU.

strictions in fact work as ARIs, as they are no longer DMU-specific.

$$\phi_r \leq \frac{u_r \ \bar{y}_r}{\sum_{r=1}^{s} u_r \ \bar{y}_r} \leq \psi_r \tag{3.15}$$

Besides the advantage of avoiding the problems associated with the DMU-specific weight restrictions previously discussed, the bounds of the restrictions become independent of the units of measurement of the outputs, as the numerator and denominator are the product of the raw weights with the output quantities. Thus, the bounds ϕ_r or ψ_r of expression (3.15) may be interpreted as the percentual importance of the output y_r in the assessment. Values of ϕ_r and ψ_r equal to one mean that the output y_r is the only one to be considered in the assessment, whereas values equal to zero mean that the corresponding output should be ignored.

The ARI weight restrictions shown in (3.15) can also be added to the CI models (3.2) and (3.8). The specification that accommodates the ARI restrictions to both desirable and undesirable outputs in model (3.2) is shown in (3.16). Note that in this case the denominator no longer coincides with the normalization constraint, so it cannot be omitted.

$$\begin{cases} \phi_r \leq \dfrac{u_r \ \bar{y}_r}{\sum_{r=1}^{s} u_r \ \bar{y}_r + \sum_{k=1}^{l} p_k \ (M_k - \bar{b}_k)} \leq \psi_r \\ \phi_k \leq \dfrac{p_k \ (M_k - \bar{b}_k)}{\sum_{r=1}^{s} u_r \ \bar{y}_r + \sum_{k=1}^{l} p_k \ (M_k - \bar{b}_k)} \leq \psi_k \end{cases} \tag{3.16}$$

Similarly, the specification of the output restrictions for the model (3.8) is shown in (3.17).

$$\begin{cases} \phi_r \leq \dfrac{u_r \ \bar{y}_r}{\sum_{r=1}^{s} u_r \ \bar{y}_r + \sum_{k=1}^{l} p_k \ \bar{b}_k} \leq \psi_r \\ \phi_k \leq \dfrac{p_k \ \bar{b}_k}{\sum_{r=1}^{s} u_r \ \bar{y}_r + \sum_{k=1}^{l} p_k \ \bar{b}_k} \leq \psi_k \end{cases} \tag{3.17}$$

Restrictions (3.16) and (3.17) can be also generalised to restrictions to categories of indicators ($C_z,\ z = 1, ..., q$), as shown in (3.18) and (3.19), respectively.

$$\phi_z \leq \frac{\sum_{r \in C_z} u_r\ \bar{y}_r + \sum_{k \in C_z} p_k\ (M_k - \bar{b}_k)}{\sum_{r=1}^{s} u_r\ \bar{y}_r + \sum_{k=1}^{l} p_k\ (M_k - \bar{b}_k)} \leq \psi_z \tag{3.18}$$

$$\phi_z \leq \frac{\sum_{r \in C_z} u_r\ \bar{y}_r + \sum_{k \in C_z} p_k\ \bar{b}_k}{\sum_{r=1}^{s} u_r\ \bar{y}_r + \sum_{k=1}^{l} p_k\ \bar{b}_k} \leq \psi_z \tag{3.19}$$

In the illustrative application discussed next, we demonstrate the advantages and limitations of using virtual weight restrictions and ARIs in the context of the construction of composite indicators including both desirable and undesirable outputs.

3.5 Illustrative application

This section illustrates the application of the two approaches to construct DEA-based CIs including desirable and undesirable outputs, presented in section 3.3. It also discusses the implications of using weight restrictions in this context.

Our illustrative example consists of a set of 9 DMUs. To allow a graphical illustration of the models, these DMUs are assessed considering two output indicators: Y, a desirable output, and B, an undesirable output. Table 3.1 shows the data for the 9 DMUs.

Table 3.1: Output indicators for the illustrative example

DMU	Y (desirable)	B (undesirable)
A	5	7
B	25	28
C	22	15
D	10	18
E	30	30
F	20	22
G	19.80	17.80
H	20.80	19.50
I	19.08	19.66

3.5.1 CI models with desirable and undesirable outputs

3.5.1.1 Indirect approach illustration

Figure 3.1 illustrates the production possibility set for the illustrative example presented in Table 3.1, corresponding to an evaluation using model (3.2). As lower values of the output indicator B correspond to better performance, the technically efficient frontier is given by the segments linking DMUs A, C and E.

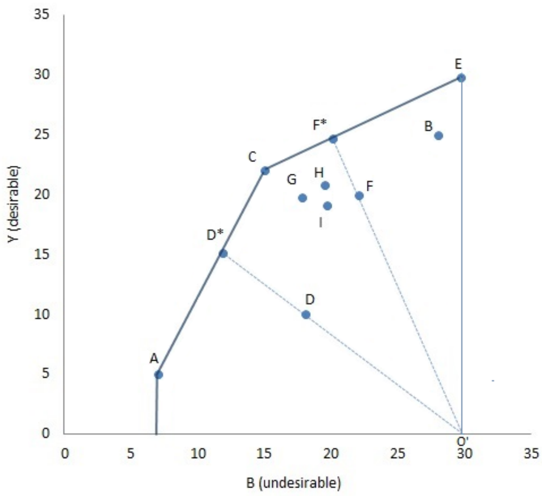


Figure 3.1: Production possibility set for the indirect approach

Table 3.2 shows the performance scores, peers and targets obtained for the 9 DMUs using model (3.2) and its dual (3.3). The value of M was set to be equal to 30 (i.e., the largest value observed for the undesirable output indicator B). The composite indicator (or efficiency score) for a DMU is given by $1/h_{j_0}$ at the optimal solution to model (3.2). For example, for DMU F the CI is 0.809, which corresponds to the ratio $\frac{O'F}{O'F^*}$ in Figure 3.1. The target that the DMU F should achieve to improve performance and reach the frontier of the production possibility set is given by the point F*, which corresponds to the value 24.725 for the output indicator Y and 20.110 for the output indicator B. The peers for DMU F are DMUs C and E, with values of λ_C and λ_E equal to 0.659 and 0.341, respectively. The values of λ_j provide an indication of the degree of similarity between a DMU and its peers.

Table 3.2: Composite indicator, peers and targets obtained using model (3.2), with M equal to 30

DMU	CI ($1/h_{j_0}$)	Peers (λ)	Target for Y	Target for B
A	1	A(1)	5	7
B	0.869	C (0.153); E (0.847)	28.772	27.698
C	1	C(1)	22	15
D	0.659	A (0.401); C (0.599)	15.176	11.789
E	1	E(1)	30	30
F	0.809	C (0.659); E (0.341)	24.725	20.110
G	0.877	C (0.928); E (0.072)	22.580	16.087
H	0.880	C (0.795); E (0.205)	23.636	18.068
I	0.820	C (0.841); E (0.159)	23.273	17.387

With the transformation in the measurement scale of the undesirable output, the computation of the composite indicator uses as basis a point that no longer coincides with the origin. Instead, it corresponds to a new reference point that depends on the value of the constant M used in model (3.2). In our illustrative example, the composite indicator is estimated considering that the projection starts at the worst value observed in the original measurement scale of the undesirable output (30, corresponding to DMU E), as shown in Figure 3.1.

As mentioned in section 3.3, the transformation consisting on subtracting the values of undesirable outputs from a large positive number (M) has an impact on the results of the DEA model. Table 3.3 presents the results that would be obtained by specifying different values for M. Note that as the value of M increases, the discrimination between the DMUs' scores decreases. In addition, the improvement required for the DMUs to reach the frontier becomes more demanding for the undesirable output and less demanding for the desirable output. This effect can be seen in Figure 3.2, that shows the projections obtained using M equal to 35 and 60.

Table 3.3: Composite indicator and ranks for different values of M

DMU	$M = 30$		$M = 35$		$M = 60$	
	CI	Rank	CI	Rank	CI	Rank
A	1	1	1	1	1	1
B	0.869	6	0.880	6	0.914	6
C	1	1	1	1	1	1
D	0.659	9	0.715	9	0.844	9
E	1	1	1	1	1	1
F	0.809	8	0.824	8	0.875	8
G	0.877	5	0.887	5	0.931	4
H	0.880	4	0.890	4	0.928	5
I	0.820	7	0.834	7	0.891	7

By using a value of M larger than the maximum observed for the undesirable output, the DMUs' classification, as efficient or inefficient, remains unchanged, but the DMUs' efficiency score changes, and the ranking of the DMUs may also be different. For example, Figure 3.2 shows that when the constant M is specified as equal to 60, DMU G is projected to the segment linking DMUs A and C instead of the segment linking C and E, where it was projected when M was specified as equal to 35. This led to a change in the efficiency ranking of DMUs, as shown in Table 3.3. Thus, using the indirect approach to construct a composite indicator, it is only possible to ensure that the assessment is classification invariant for different values of the constant M, as stated by Seiford and Zhu (2002).

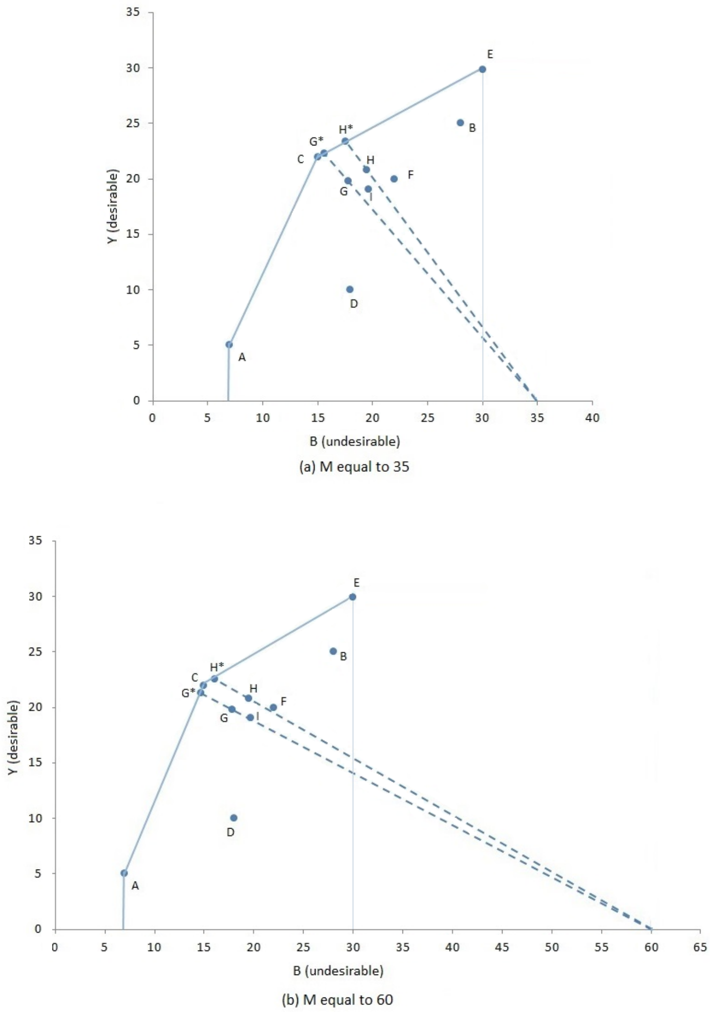


Figure 3.2: Sensibility analysis for different values of M

3.5.1.2 Direct approach illustration

This section illustrates the estimation of composite indicators using the Directional CI model. Figure 3.3 shows the production frontier that would be obtained for our illustrative example (Table 3.1) using the Directional CI model. The efficient frontier is defined by the segments linking O, C and E.

By setting the directional vector as $g = (g_y, -g_b) = (y_{rj_0}, -b_{tj_0})$, i.e. the current value of the outputs for the DMU under assessment, it is possible to simultaneously expand the desirable outputs and contract the undesirable outputs through a path that allows proportional interpretation of improvements. In order to facilitate the interpretation of the DMUs' projection to the frontier, Figure 3.3 illustrates the directional vectors and the projection of DMUs D and F on the frontier, corresponding to points D* and F*, respectively. Note that, for each DMU, the desirable and undesirable outputs are expanded and contracted, respectively, according to a direction that corresponds to proportional changes to the original levels.

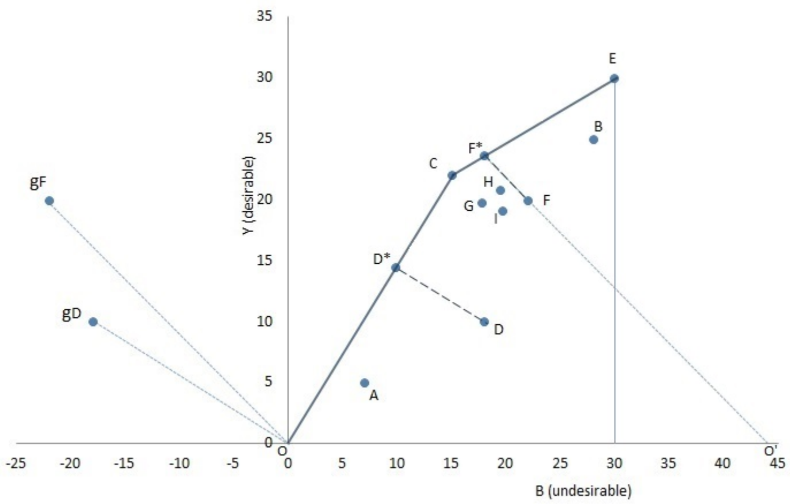


Figure 3.3: Production possibility set for the direct approach

Table 3.4 shows the composite indicator, peers and targets obtained using the Directional CI model (3.6). The value of β^* obtained at the optimal solution to the model can be interpreted as the scope for improvement of a given DMU. For example, the value of β^* for DMU F is 0.181, corresponding to $\frac{FF^*}{O'F}$ in Figure 3.3. The point F* is the target that the DMU F should achieve to become efficient, i.e., to operate at the frontier, which corresponds to the value 23.613 for the output indicator Y and 18.025 for the output indicator B. The peers for DMU F are DMUs C and E, with values of λ_C and λ_E equal to 0.798 and 0.202, respectively.

Table 3.4: Composite indicator, rank, peers and targets obtained from the Directional CI model

DMU	β	CI	Rank	Peers (λ)	Target for Y	Target for B
A	0.345	0.744	8	C (0.306)	6.725	4.585
B	0.098	0.910	3	C (0.317); E (0.683)	27.462	25.242
C	0	1	1	C(1)	22	15
D	0.451	0.689	9	C (0.659)	14.505	9.890
E	0	1	1	E(1)	30	30
F	0.181	0.847	6	C (0.798); E (0.202)	23.613	18.025
G	0.126	0.888	5	C (0.963); E (0.037)	22.296	15.556
H	0.115	0.897	4	C (0.850); E (0.150)	23.200	17.250
I	0.183	0.845	7	C (0.929); E (0.071)	22.567	16.063

As explained in section 3.3.2, from the factor β^* it is possible to obtain an efficiency measure, given by $1/(1+\beta^*)$, corresponding to the best level of performance observed. Thus, the Directional CI score for the DMU F would be equal to 0.847 $(= 1/(1+0.181))$, which corresponds to the ratio $\frac{O'F}{O'F^*}$ in Figure 3.3.

If instead of assuming weak disposability of undesirable outputs, it was imposed strong disposability, the frontier would be given by the horizontal extension of E to the Axis Y, as shown in Figure 3.4. Under this condition, the DMU F, for example, would have a score $\beta = 0.5$ and it would be projected to the point F* on the frontier with coordinates (30, 11). This means that, in order to be producing on the frontier, DMU F should reduce by half the production of undesirable outputs whilst keeping the same production level of desirable outputs. Note that depending on the directional vector used, the projection of some DMUs to the frontier could correspond to negative values of undesirable outputs. This would be the case of DMU D, whose score under strong disposability of the undesirable output is equal to $\beta = 2$. This requires an unreasonable improvement for the undesirable output, since the DMU would be projected to point D* (30, -18).

Therefore, in the context of assessments using a CI, we believe that weak disposability is more appropriate, since it avoids unrealistic projections, i.e., projections involving huge reductions to undesirable outputs and increments to desirable outputs, eventually leading to unrealistic targets corresponding to negative values of the undesirable outputs.

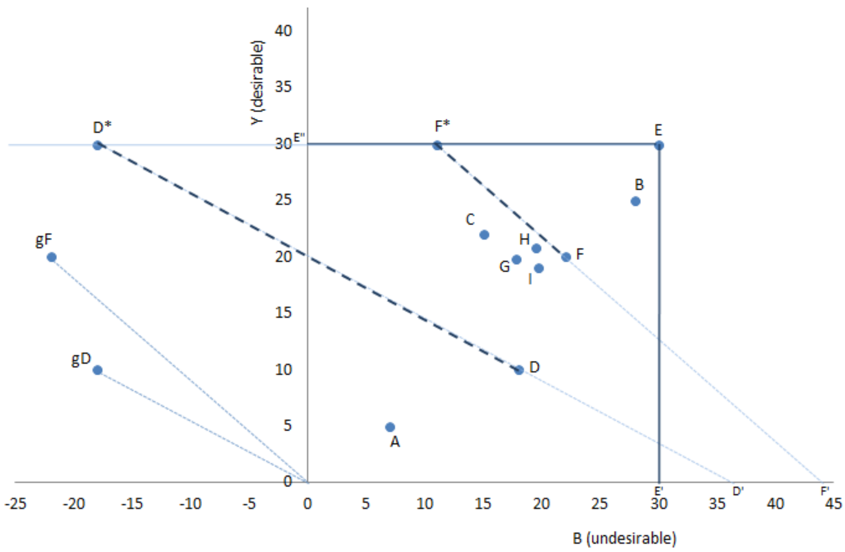


Figure 3.4: Production possibility set for the direct approach assuming strong disposability of undesirable outputs

3.5.1.3 Discussion regarding the alternative formulations of the CI model

As shown in the illustrative example, the efficient frontiers obtained using the direct and indirect approaches are different (see Figures 3.1 and 3.3). The weak disposability assumption imposed to the Directional CI model forces the frontier to pass through the origin. Conversely, the assessment using model (3.2) results in a frontier that does not pass through the origin. This implies that the classification of a DMU as efficient or inefficient depends on the model used. Furthermore, the performance scores and ranking of DMUs obtained with the direct and indirect approaches are also different.

The advantages of using the Directional CI model, shown in (3.6), instead of the CI model, shown in (3.2), is that it allows to estimate, simultaneously, the inefficiencies associated with desirable and undesirable outputs without any transformation in the measurement scale of undesirable outputs. Furthermore, by setting the components of the directional vector equal to the current value of the outputs for the DMU under assessment, it is possible to preserve the proportional interpretability of the improvements. Conversely, as the CI specified in (3.2), involving a change in the measurement scale of the undesirable outputs, computes all projections in relation to a fixed point that does not coincide with the origin, it does not allow a proportional change to both desirable and undesirable outputs, and the CI score depends on the values chosen for the constant M_k. Thus, we conclude that the Directional CI model is the best alternative for assessments involving both desirable and undesirable outputs.

3.5.2 Estimation of CI with weight restrictions

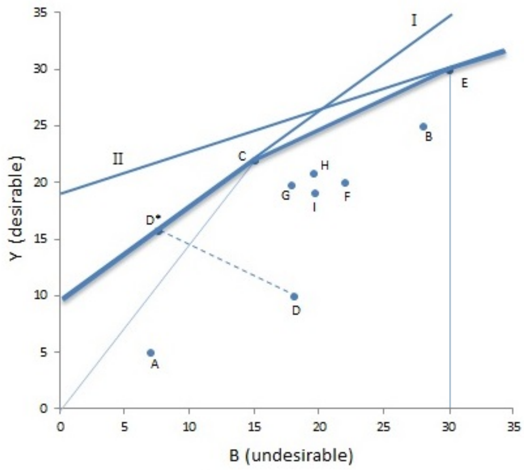
This section discusses the use of restrictions to virtual outputs and ARI restrictions in the context of assessments involving composite indicators with both desirable and undesirable outputs. These restrictions are specified in order to reflect the relative importance of the indicators in percentage terms. We only illustrate these restrictions in the context of the Directional CI model, given its advantages for assessments involving CI, as previously discussed. The weight restrictions are imposed to the dual of model (3.6), shown in formulation (3.8).

We start by discussing the use of virtual weight restrictions presented in (3.14). In order to illustrate the impact of these restrictions in the estimation of the Directional CI model (3.8) for the 9 DMUs presented in Table 3.1, consider that the virtual weight of each output should be at least 40% of the total virtual weights, as shown in (3.20).

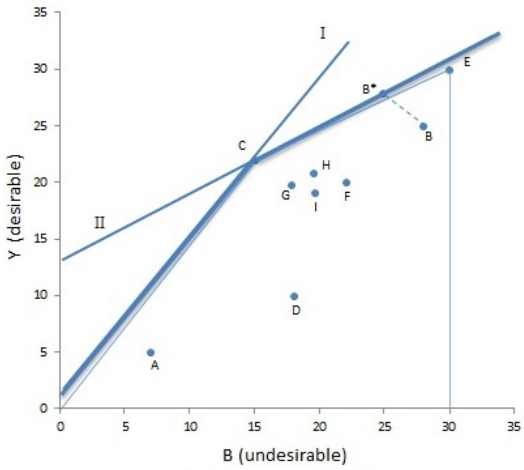
$$\begin{cases} u_1 \; y_{1j_0} \geq 0.4 \\ p_1 \; b_{1j_0} \geq 0.4 \end{cases} \tag{3.20}$$

The changes to the efficient frontier resulting from the use of weight restrictions (3.20) with model (3.8) for the evaluation of DMUs D and B, are shown in Figure 3.5 (a) and (b), respectively.

Considering the assessment of DMU D, the weight restriction imposed to the desirable output ($u_1 \; 10 \geq 0.4$) determines the segment labelled I, whereas segment II represents the restriction associated with the undesirable output ($p_1 \; 18 \geq 0.4$). This implies that the slope of the frontier must be between the slopes of segments I and II to ensure that for both outputs are given a weight representing at least 40% of the total virtual weight of DMU D. This leads to an evaluation against the frontier defined by the segments in bold. Note that the segment of the original frontier linking the origin to DMU C corresponded to an assessment where the virtual weight given to the desirable output was smaller than 40%, and thus this segment can no longer be used for the estimation of the CI of DMU D.



(a) Assessment of DMU D



(b) Assessment of DMU B

Figure 3.5: Assessment of DMUs D and B with the DMU-specific virtual weight restrictions added to the Directional CI model

Regarding the assessment of DMU B, it can be seen that the frontier against which this DMU is assessed is different from the frontier underlying the evaluation of DMU D. This happens because the virtual weight restrictions are DMU-specific and thus different restrictions lead to the specification of different frontiers. The targets that the DMUs should pursue in order to reach the frontier are given by (y=15.8, b=7.56) for DMU D and (y=27.84, b=24.82) for DMU B.

As the CI is estimated based on comparisons with different frontiers, one for each DMU assessed, some DMUs considered inefficient when evaluated against their specific frontier may appear as peers for others, as they can be located on the frontier corresponding to the use of a different set of weights. Note that the weights imposed in their own evaluation may be more restrictive that the weights allowed using a different specification of the virtual weight restrictions, corresponding to the assessment of other DMUs.

Next we illustrate the use of ARI weight restrictions presented in (3.17). For the example described in Table 3.1, consider once again that we want to ensure that the virtual weight of each output indicator should be at least 40% of the total virtual weight. The corresponding weight restrictions to be imposed to the model (3.8) are shown in (3.21).

$$\begin{cases} \dfrac{u_1\ \bar{y}_1}{u_1\ \bar{y}_1 + p_1\ \bar{b}_1} \geq 0.4 \\ \dfrac{p_1\ \bar{b}_1}{u_1\ \bar{y}_1 + p_1\ \bar{b}_1} \geq 0.4 \end{cases} \tag{3.21}$$

In Figure 3.6, the restrictions for the desirable and undesirable outputs are represented by the segments I and II, respectively. The slope of the efficient frontier must be between the slopes of these segments to ensure that both outputs are weighted at least 40% in the assessment. The efficient frontier used for the assessment of all DMUs is defined by the segments in bold. Figure 3.6 also illustrates the projection of the DMUs on the frontier. For example, the targets that DMUs D and B should pursue in order to become efficient are given by ($y = 14.61, b = 9.86$) and ($y = 28.14, b = 24.49$), respectively.

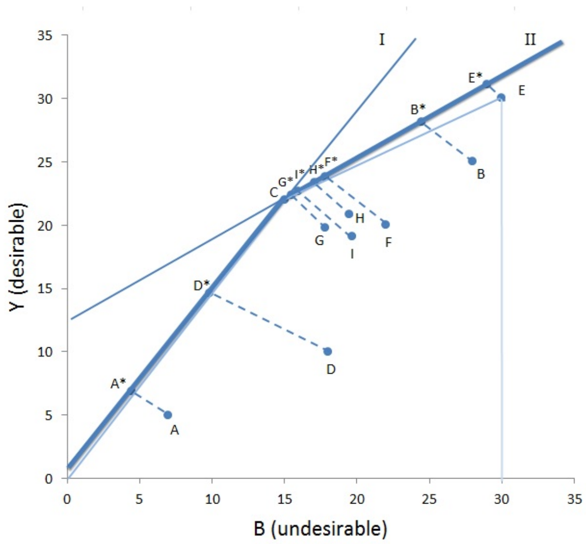


Figure 3.6: Assessment of all DMUs with the ARI weight restrictions added to the Directional CI model

Unlike what happened with the virtual weight restrictions, which led to assessments against DMU-specific frontiers, using the ARI type of restrictions all DMUs are assessed against a unique frontier, as the weight restrictions are identical for all DMUs.

Table 3.5 shows the scores, rank and peers obtained using the formulations of the Directional CI model with different types of restrictions. The first column reports the results of the unrestricted model, and the second and third columns show the results of the Directional CI model with virtual weight restrictions and with ARI restrictions, respectively.

By comparing the results obtained using the ARI weight restrictions and virtual weight restrictions, we can see that the model with the ARI provided results more similar to those of the unconstrained model. This could be expected because the virtual weight restrictions lead to the construction of DMU-specific frontiers, and this variability in the shape of the reference frontier leads to efficiency scores that differ considerably from the original values.

Table 3.5: Comparison of the composite indicator scores of the Directional CI models, with and without weight restrictions

	Original model	Virtual weight restrictions	ARI weight restrictions
DMU	CI	CI	CI
A	0.744	0.597	0.739
B	0.910	0.898	0.889
C	1	1	1
D	0.689	0.633	0.689
E	1	0.9622	0.967
F	0.847	0.842	0.840
G	0.888	0.885	0.886
H	0.897	0.887	0.890
I	0.845	0.843	0.843

It is also worth noting that only one DMU (DMU C) had the same efficiency score in all models. This is because its weighting system in the original evaluation, without weight restrictions, was within the bounds specified for both models with virtual weight restrictions and with ARI restrictions. In general, when the original weighting system is similar to the balance imposed by the bounds to the weights, the results of the restricted models tend to be closer to the original values. Conversely, when the original weighting system is quite different from the bounds imposed by the models with weight restrictions, the differences in the scores obtained from different models can be large.

In the illustrative example, the outputs of DMU I are equal to the sample average for each output. In this particular case, the virtual weight restrictions specified for this DMU are identical to the ARI restrictions based on the use of the sample average. Therefore, the results of both models are the same.

As the virtual weight restrictions are not able to construct a unique frontier for all DMUs analysed, they lead to comparisons based on different feasible regions. For this reason, we believe they are not the best option to construct composite indicators and ranks. A fair comparison requires the DMUs to be assessed based on similar conditions (Ramon et al., 2011; Hatefi and Torabi, 2010). The fact that the ARI are more conservative than virtual weight restrictions, combined with the fact that the virtual weight restrictions can suggest peer DMUs that are inefficient when evaluated with their own weight restrictions, led us to conclude that the approach based on ARI restrictions is the most appropriate.

3.6 Conclusions

The traditional DEA-based composite indicator models cannot be used in the presence of both desirable and undesirable outputs, as they do not seek for reductions to the undesirable indicators. In this chapter we discussed two different approaches that can be used for the construction of CI in this context: an indirect approach, based on a traditional DEA model including a transformation in the measurement scale of undesirable outputs, and a direct approach, based on a DEA model specified with a directional distance function, that allows dealing with the undesirable outputs in their original measurement scale.

In order to explain the features of the approaches discussed in this chapter we illustrated their implementation using a small example. It was demonstrated that in the indirect approach, after the transformation in the measurement scale of the undesirable outputs, the reference point used to compute the measure of performance is no longer the origin, as in standard DEA models, but a new reference point whose coordinates are equal to the positive numbers used to transform the measurement scale of the undesirable outputs. As a result, this approach does not allow proportional improvements to both desirable and undesirable outputs. Furthermore, it was shown that the results of the indirect approach are sensitive to the value of the constant used for the transformation of the measurement scale. Different values for the constant have implications both in the performance scores and ranking of the DMUs.

Conversely, using the direct approach to deal with the undesirable outputs, it is possible to preserve the proportional interpretability of the improvements by setting the components of the directional vector equal to the values of the desirable and undesirable outputs of the DMU under assessment. This is an important advantage of the direct approach, and thus we argue that it is the most appropriate for constructing composite indicators in the presence of both desirable and undesirable outputs.

This chapter also explored different ways to incorporate information on decision-maker preferences about the relative importance of individual indicators aggregated in the CI. The specification of two different types of weight restrictions that can be used in this context (virtual weight restrictions and ARI weight restrictions) was discussed and illustrated using a small example. We suggested an enhanced specification of the ARI weight restrictions that allows incorporating the relative importance of outputs, expressed as a percentage. This formulation also has the advantage of being independent of the units of measurement of the output indicators.

The ARI weight restrictions overcome some limitations of virtual weight restrictions that are commonly used for evaluations based on CI. The problem of having peers for inefficient DMUs that are not efficient when assessed with their own set of virtual weights does not occur with the specification of the enhanced ARI restrictions, as they avoid evaluations against different frontiers. Thus, we also conclude that the new restrictions proposed here, in the form of ARI, are the most appropriate approach to reflect the relative importance of outputs in assessments involving the use of CI.

The specification of this novel type of weight restriction and the construction of DEA-based CIs that can accommodate both desirable and undesirable outputs using a directional distance function are the two major methodological contributions of this chapter.

Chapter 4

An enhanced Malmquist-Luenberger index to assess productivity change in the presence of undesirable outputs

4.1 Introduction

The Malmquist productivity index, introduced by Caves et al. (1982) and then developed by Fare et al. (1994b), is the most frequently used approach to assess productivity change over time. As the Malmquist productivity index is calculated based on ratios of Shephard's distance functions, it can either have an input or output orientation. As the distance functions cannot consider simultaneous adjustments to inputs and outputs or reductions to a subset of outputs and increment to other outputs, the Malmquist index cannot accommodate assessments with undesirable outputs.

In order to overcome these limitations, Chung et al. (1997) proposed an adaptation to the Malmquist productivity index that can accommodate undesirable outputs. The new index, named Malmquist-Luenberger (ML) index, allows DMUs to pursue simultaneous contractions of inputs and undesirable outputs and expansion of desirable outputs. Instead of estimating Shephard's distance functions, the ML index is calculated using ratios of directional distance functions. This approach has been extensively applied in the literature to measure changes in productivity over time in the presence of undesirable outputs.

A different approach that can also overcome the limitations of the Malmquist index was proposed by Chambers (1996). The authors introduced the Luenberger productivity index to assess productivity change over time using the difference of directional distance functions.

While the Malmquist index focuses on either inputs contraction or outputs expansion (cost minimization or revenue maximization), the Malmquist-Luenberger index and the Luenberger index can consider simultaneously inputs contraction and outputs expansion (profit maximization). Although these indices were designed to account for simultaneous improvements in inputs and outputs, they can also be specified with an output or input orientation when necessary. In this sense, we can say that the Malmquist-Luenberger index and the Luenberger index encompass the Malmquist index.

Boussemart et al. (2003) pointed out the three main differences between the Malmquist and Luenberger indices, which can motivate the use of one or another. The first is related to the choice of the distance function (Shephard or the directional distance functions). The second is associated with the economic motivation, which can focus on revenue/cost optimization or on profit maximization. The last difference is related to the nature of the index (multiplicative or additive). The first two issues mentioned by Boussemart et al. (2003) can also differentiate the Malmquist from the Malmquist-Luenberger index. Both the Malmquist-Luenberger and Luenberger indices use the directional distance function to estimate the index, and thus, can account for revenue/cost optimization or profit maximization. The main methodological difference between the Malmquist-Luenberger and Luenberger indices is related to the multiplicative or additive nature.

Several empirical applications used the ratio-based Malmquist-Luenberger index to assess productivity change (e.g. Zhang et al. (2011); Krautzberger and Wetzel (2012); He et al. (2013)). On the other hand, fewer empirical applications used the Luenberger productivity index, which measures productivity change in terms of differences rather than ratios (e.g. Epure et al. (2011); Briec et al. (2011); Williams et al. (2011)). As noted by Epure et al. (2011), although the ratio-based indices (such as Malmquist and Malmquist-Luenberger indices) are more familiar in the academic community, in the business and accounting communities the difference-based indices may be more obvious, since they measure cost, revenue, or profit differences in monetary terms.

This chapter addresses the different approaches that can be used to accommodate undesirable outputs in the analysis of productivity change over time. We start from the approach proposed by Chung et al. (1997), in which the ML index is derived from a standard Malmquist index using the relationship between the directional distance function and the Shephard's output distance function. Next, we show that an equivalent index can be derived using the relationship between the directional distance function and Shephard's input distance function. In the context of assessments involving both desirable and undesirable outputs, the two versions of the ML index represent equally good adaptations of the Malmquist index. As the indices provide different results, we propose the use of an enhanced version of the ML index, given by the geometric mean of the two former versions of the ML indices. This approach avoids an arbitrary selection of the input or output Shephard's distance function to derive the Malmquist-Luenberger index. In assessments in which improvements in both directions are required (reducing undesirable outputs and increasing desirable outputs), the Average ML proposed in this chapter has the advantage of representing more accurately the changes in DMUs' productivity over time, as it incorporates both orientations in the computation of the productivity change score.

We used a case study to compare the results obtained by the different versions of the ML indices with the Luenberger index, which is an established difference-based measure to assess productivity change over time considering simultaneous adjustments to desirable and undesirable outputs. The different indices are applied to evaluate the performance change over time of the European Commercial Transport Industry. The data concern 17 European countries in years 2005 and 2006. Based on the empirical results, we explore the relationship between the Malmquist-Luenberger index proposed in this chapter and the Luenberger index, both in terms of the values of productivity change estimates and the rankings obtained. It is shown that the Average ML index proposed in this chapter is a robust alternative to the Luenberger index, which can be used when a multiplicative index based on ratios is considered preferable to a difference based index. The limitations of the input and output ML index are highlighted by comparison to the Average ML index.

The generalisation of the enhanced ML index proposed in this section to assessments involving composite indicators is straightforward, as it only requires replacing the directional distance function used for the estimation of the ML index by the CI model based on the directional distance function described in the previous section.

The remainder of this chapter proceeds as follows. Section 4.2 approaches the measurement of productivity change over time using the Malmquist index. Section 4.3 presents the main nonparametric indices that can be used to measure productivity change over time in the presence of undesirable outputs. These two sections constitute a literature review that provide the foundations for the enhanced Malmquist-Luenberger index proposed in section 4.4. Section 4.5 presents a graphical illustration of the productivity change indices and explores, using a real-world application, the robustness of the different productivity indices approached in this chapter. Finally, section 4.6 concludes the chapter.

4.2 Productivity change over time

The Malmquist index is considered in the literature the standard approach to evaluate productivity change over time. Before showing its formulation, we start by presenting the concept of a distance function for a technology involving the use of multiple inputs to produce multiple outputs, which underlies the construction of the index.

4.2.1 Shephard's distance functions

Consider that the production technology T models the transformation of inputs, denoted by $x \in \Re^m_+$, into outputs, denoted by $y \in \Re^s_+$, as shown in (4.1). The production technology consists of the set of all feasible input/output vectors for a certain production process.

$$T = \{(x, y) : x \text{ can produce } y\} \tag{4.1}$$

Following Shephard (1970) and Fare et al. (1994a), the output distance function for DMU j_0 in relation to the technology T is defined as shown in (4.2).

$$D_o(x,y) = \min\{\theta : (x, \frac{y}{\theta}) \in T\} \tag{4.2}$$

This function gives the reciprocal of the maximum factor $1/\theta$ by which the output vector y can be proportionally expanded, given inputs x. This means that it corresponds to the efficiency score of DMU j_0, i.e. $D_o(x,y) \leq 1$. Note that $D_o(x,y) \leq 1$ if and only if $(x,y) \in T$. In particular, $D_o(x,y) = 1$ if and only if (x,y) is on the frontier of the technology, meaning that the production point corresponding to (x,y) is technically efficient.

Assuming constant returns to scale, the efficiency of the DMU j_0 can be determined using the linear programming problem shown in (4.3), proposed by Charnes et al. (1978). Thus, the distance function can be estimated using a Data Envelopment Analysis (DEA) model. Fare et al. (1994b) were the first to note that input and output distance functions could be estimated using DEA models.

$$\begin{aligned}
&(D_o(x,y))^{-1} = \text{Max } \theta & & \qquad (4.3)\\
&s.t. \sum_{j=1}^{n} y_{rj} \; \lambda_j \geq \theta y_{rj_0} & & r = 1, \ldots, s \\
&\sum_{j=1}^{n} x_{ij} \; \lambda_j \leq x_{ij_0} & & i = 1, \ldots, m \\
&\lambda_j \geq 0 & & j = 1, \ldots, n
\end{aligned}$$

The λ_j are the intensity variables. The factor $1/\theta$ indicates the DMU's efficiency.

Similarly, the input distance function is defined as shown in (4.4)[1].

[1] A more rigorous definition could be made by replacing *min* and *max* (which stands for *minimum* and *maximum*) with *inf* and *sup* (which stands for *infimum* and *supremum*), because the *min* and *max* may not be attained. However, in the interests of easy reading, the terms *min* and *max* are frequently used (see Coelli et al. (2005, p.49) and Fried et al. (2008, p.22)).

$$D_i(x,y) = \max\{\delta : (\frac{x}{\delta}, y) \in T\} \tag{4.4}$$

The input distance function gives the reciprocal of the minimum factor $1/\delta$ by which the input vector x can be proportionally contracted, given outputs y. The input technical efficiency is therefore defined as $1/D_i(x,y)$. For the DMU j_0, the input oriented efficiency can be obtained through the linear programming problem shown in (4.5), proposed by Charnes et al. (1978). This DEA model also assumes constant returns to scale.

$$\begin{aligned}
&(D_i(x,y))^{-1} = \text{Min } \delta && \\
&s.t. \sum_{j=1}^{n} y_{rj}\ \lambda_j \geq y_{rj_0} && r = 1,\ldots,s \\
&\sum_{j=1}^{n} x_{ij}\ \lambda_j \leq \delta x_{ij_0} && i = 1,\ldots,m \\
&\lambda_j \geq 0 && j = 1,\ldots,n
\end{aligned} \tag{4.5}$$

Under constant returns to scale, the following relationship holds for the distance functions: $D_o(x,y) = (D_i(x,y))^{-1}$.

4.2.2 Directional distance functions

Chambers et al. (1996), based on Luenberger (1992a,b) shortage function, proposed a directional distance function that allows a producer to scale input and outputs simultaneously along a path that is defined according to a directional vector g. The general form of the directional distance function is presented in (4.6).

$$\vec{D}(x,y;g_x,g_y) = \max\ \{\beta : (x + \beta g_x, y + \beta g_y) \in T\} \tag{4.6}$$

The components of the nonzero vector g indicates the direction of change for the inputs and outputs. When the components of the directional vector are set as $g = (-x_{ij_0}, y_{rj_0})$, i.e. the current value of the outputs for the DMU under assessment, it is possible to scale inputs and outputs through a path that allows proportional interpretation of improvements.

The directional distance function (4.6) can be solved by the linear programming problem shown in (4.7), that assumes constant returns to scale (Chambers et al., 1996).

$$\begin{aligned}
&\vec{D}(x, y; g_x, g_y) = \max \beta && \\
&s.t. \sum_{j=1}^{n} y_{rj}\ \lambda_j \geq y_{rj_0} + \beta g_{y_{rj_0}} && r = 1, \ldots, s \\
&\sum_{j=1}^{n} x_{ij}\ \lambda_j \leq x_{ij_0} - \beta g_{x_{ij_0}} && i = 1, \ldots, m \\
&\lambda_j \geq 0 && j = 1, \ldots, n
\end{aligned} \tag{4.7}$$

The factor β indicates the extend of DMU's inefficiency. It corresponds to the maximal feasible contraction of inputs and expansion of outputs that can be achieved simultaneously. Thus, when $(x, y) \in T$, $\vec{D}(x, y; g_x, g_y) \geq 0$, and $\vec{D}(x, y; g_x, g_y) = 0$ only when DMU j_0 is operating on the frontier of the technology.

4.2.3 Relation between the Shephard's distance functions and the directional distance function

Fare and Grosskopf (2000) presented the relationship between the Shephard's input and output distance functions and the directional distance functions for an example describing the transformation of inputs x_{ij} $(i = 1, ..., m)$ into outputs y_{rj} $(r = 1, ..., s)$. If the vector of the directional distance function is defined as equal to $g = (0, y_{rj_0})$, the relationship can be written as

$$\vec{D}(x, y; 0, y) = \frac{1}{D_o(x, y)} - 1 \tag{4.8}$$

By rearranging the terms in expression (4.8), we obtain an equivalent expression as shown in (4.9).

$$D_o(x, y) = \frac{1}{1 + \vec{D}(x, y; 0, y)} \tag{4.9}$$

Therefore, when the directional vector is specified as the current value of the outputs of the DMUs, the Shephard's output distance function is a special case of the directional distance function. Figure 4.1 facilitates the interpretation of the relationship between the distance functions. While the inefficiency measure, given by the directional distance function $\vec{D}(x, y; 0, y)$, for the DMU A corresponds to the ratio $\frac{AA^*}{AA'}$, the efficiency measure, given by the distance function $D_o(x, y)$, corresponds to the ratio $\frac{AA'}{A^*A'}$.

The same idea works for a vector specified as $g = (-x_{ij_0}, 0)$ in the directional distance function. This vector allows specifying a direct relationship between the directional distance function and the Shephard's input distance function, which can be written as follows:

$$\vec{D}(x, y; -x, 0) = 1 - \frac{1}{D_i(x, y)} \tag{4.10}$$

or equivalently,

$$D_i(x, y) = \frac{1}{1 - \vec{D}(x, y; -x, 0)} \tag{4.11}$$

Therefore, when the directional vector is specified as the current value of the inputs of the DMUs, the Shephard's input distance function is also a special case of the directional distance function. In this case, the inefficiency measure for the DMU A, given by the directional distance function $\vec{D}(x, y; -x, 0)$, corresponds to the ratio $\frac{AA^{**}}{AA''}$ in Figure 4.1, and the efficiency measure, given by the inverse of the input distance function $(1/D_i(x, y))$, is equal to the ratio $\frac{A^{**}A''}{AA''}$.

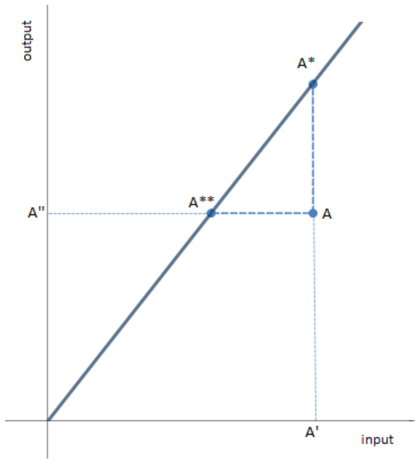


Figure 4.1: Input and output distance functions

4.2.4 The Malmquist index

The Malmquist index (MI), introduced by Caves et al. (1982) and developed by Fare et al. (1994b), is defined as a ratio of input (or output) distance functions applied to the assessment of productivity change over time. The index was named after Sten Malmquist, who in 1953 defined input quantity indices as ratios of distance functions. It is the most frequently used approach to assess productivity change over time.

The output oriented Malmquist index requires the specification of Shephard's output distance functions for the time periods t and $t+1$, $D_o^t(x^t, y^t)$ and $D_o^{t+1}(x^{t+1}, y^{t+1})$, respectively, as well as two additional distance functions:

$$D_o^t(x^{t+1}, y^{t+1}) = \min\{\theta : (x^{t+1}, \frac{y^{t+1}}{\theta}) \in T^t\} \tag{4.12}$$

$$D_o^{t+1}(x^t, y^t) = \min\{\theta : (x^t, \frac{y^t}{\theta}) \in T^{t+1}\} \tag{4.13}$$

The superscript of the output distance functions D_o^t and D_o^{t+1} indicates the time period used to construct the reference technology, and the time period of the data being evaluated is included inside the parentheses. While in the within-period assessment, $D_o^t(x^t, y^t)$ and $D_o^{t+1}(x^{t+1}, y^{t+1})$, the value of the function is always smaller than or equal to one, in the mixed-period cases, $D_o^t(x^{t+1}, y^{t+1})$ and $D_o^{t+1}(x^t, y^t)$, the value can be smaller, equal or greater than one. Values greater than one signal cases in which the production of DMUs from a given time period occurs outside the technology of the period considered to construct the frontier.

The output oriented Malmquist index for an assessment involving inputs and outputs is defined as follows (Fare et al., 1994b):

$$MI_o^{t,t+1} = \left[\frac{D_o^t(x^{t+1},y^{t+1})}{D_o^t(x^t,y^t)} \frac{D_o^{t+1}(x^{t+1},y^{t+1})}{D_o^{t+1}(x^t,y^t,)}\right]^{\frac{1}{2}} \tag{4.14}$$

In order to avoid an arbitrary choice between the base periods, this index corresponds to a geometric mean of the periods t and $t+1$ Malmquist indices, corresponding to the two terms inside brackets.

Fare et al. (1994b) also showed how to decompose the Malmquist index in two measures, one measuring the efficiency change (EC), and one reflecting the technological change (TC), i.e. the change in the frontier of the production possibility set. These components are obtained by rewriting the Malmquist index (4.14) as follows:

$$EC_o^{t,t+1} = \frac{D_o^{t+1}(x^{t+1},y^{t+1})}{D_o^t(x^t,y^t)} \tag{4.15}$$

and

$$TC_o^{t,t+1} = \left[\frac{D_o^t(x^t,y^t)}{D_o^{t+1}(x^t,y^t)} \frac{D_o^t(x^{t+1},y^{t+1})}{D_o^{t+1}(x^{t+1},y^{t+1})}\right]^{\frac{1}{2}} \tag{4.16}$$

The product of the two components results in the Malmquist index:

$$MI_o^{t,t+1} = EC_t^{t+1} \cdot TC_t^{t+1} \tag{4.17}$$

Similarly, the input oriented Malmquist index for an assessment involving inputs and outputs is defined as shown in (4.18).

$$MI_i^{t,t+1} = \left[\frac{D_i^t(x^t,y^t)}{D_i^t(x^{t+1},y^{t+1})} \frac{D_i^{t+1}(x^t,y^t)}{D_i^{t+1}(x^{t+1},y^{t+1})}\right]^{\frac{1}{2}} \tag{4.18}$$

For the input oriented Malmquist index, the components are obtained by rewriting the index (4.18) as follows:

$$EC_i^{t,t+1} = \frac{D_i^t(x^t,y^t)}{D_i^{t+1}(x^{t+1},y^{t+1})} \tag{4.19}$$

and

$$TC_i^{t,t+1} = \left[\frac{D_i^{t+1}(x^t,y^t)}{D_i^t(x^t,y^t)} \frac{D_i^{t+1}(x^{t+1},y^{t+1})}{D_i^t(x^{t+1},y^{t+1})}\right]^{\frac{1}{2}} \tag{4.20}$$

The values of $MI_o^{t,t+1}$, $MI_i^{t,t+1}$ and their components can be greater, equal or smaller than one, indicating, respectively, if productivity growth, stagnation or decline occurred between periods t and $t+1$, respectively. Improvements in the efficiency change are evidence of catching up to the frontier and improvements in technological change are evidence of innovation. The component related to the technological change captures the shift in technology between the two time periods.

4.3 Productivity change over time in the presence of undesirable outputs

4.3.1 Distance functions that include undesirable outputs

In order to explain the concept of a distance function for a technology with desirable and undesirable outputs, consider that the production technology models the use of inputs, denoted by $x \in \Re^m_+$, to produce desirable and undesirable outputs, denoted by $y \in \Re^s_+$ and $b \in \Re^l_+$, respectively. Then, the production technology represented by the output set $P(x)$ can be described as shown in (4.21).

$$P(x) = \{(y, b) : x \text{ can produce } (y, b)\} \tag{4.21}$$

Fare et al. (1989) modeled the idea that the undesirable outputs can not be reduced without cost by imposing the assumption that desirable and undesirable outputs (y, b) are, together, weakly disposable, as shown in (4.22). When imposing weak disposability of undesirable outputs we are assuming that they are by-products of the desirable outputs, which implies that abatement in an undesirable output is possible if accompanied by a reduction in a desirable output or by using additional resources that could have been used to increase the production of desirable outputs.

$$(y, b) \in P(x) \text{ and } 0 \leq \theta \leq 1 \text{ imply } (\theta y, \theta b) \in P(x) \tag{4.22}$$

In addition, it is assumed that the good outputs are freely disposable, i.e., they can be reduced without affecting the production of undesirable outputs. This assumption can be written as shown in (4.23).

$$(y,b) \in P(x) \text{ and } \hat{y} \leq y \text{ imply } (\hat{y},b) \in P(x) \tag{4.23}$$

Finally, it is modeled the idea that the desirable outputs are jointly produced with the undesirable outputs. This means that the only way to produce zero undesirable outputs is by producing zero desirable outputs. This idea, named null-jointness, was introduced by Shephard and Fare (1974) and it is expressed as shown in (4.24).

$$\text{if } (y,b) \in P(x) \text{ and } b=0 \text{ then } y=0 \tag{4.24}$$

As shown in Chung et al. (1997), following Shephard (1970), for assessments involving both desirable and undesirable outputs, the output distance function in relation to the production technology $P(x)$ can be defined as follows:

$$D_o(x,y,b) = \min \, \{\theta : ((y,b)/\theta) \in P(x)\} \tag{4.25}$$

However, the Shephard's output distance function shown in (4.25), seeks to increase both desirable and undesirable outputs simultaneously. One possibility to overcome this problem and credit DMUs for reductions in undesirable outputs is the specification of a directional distance function. Chung et al. (1997) extended the Chambers et al. (1996) approach, presented in section 4.2.2, to allow including undesirable outputs in the evaluation. It can be defined as shown in (4.26).

$$\vec{D}(x,y,b;g) = \max \, \{\beta : (y,b) + \beta g \in P(x)\} \tag{4.26}$$

Conversely to the distance function (4.25), the directional distance function allows to simultaneously expand the desirable outputs and contract the undesirable ones. The directional distance function (4.26) can be solved by the linear programming problem shown in (4.27). It assumes constant returns to scale and satisfies the conditions (4.22), (4.23) and (4.24).

$$\begin{aligned}
&\vec{D}(x,y,b;g) = \max \beta && \qquad (4.27)\\
&s.t. \sum_{j=1}^{n} y_{rj}\ \lambda_j \geq y_{rj_0} + \beta g_{y_{rj_0}} \qquad r = 1,\ldots,s\\
&\sum_{j=1}^{n} b_{kj}\ \lambda_j = b_{kj_0} - \beta g_{b_{kj_0}} \qquad k = 1,\ldots,l\\
&\sum_{j=1}^{n} x_{ij}\ \lambda_j \leq x_{ij_0} \qquad i = 1,\ldots,m\\
&\lambda_j \geq 0 \qquad j = 1,\ldots,n
\end{aligned}$$

The factor β corresponds to the maximal feasible expansion of desirable outputs and contraction of undesirable outputs that can be achieved simultaneously.

Figure 4.2 shows an illustration of the production frontier that would be obtained using model (4.27) with the components of the directional vector set as $g = (y_{rj_0}, -b_{kj_0})$, i.e. the current value of the outputs for the DMU under assessment. As explained in section 3.3.2, by specifying $g = (y_{rj_0}, -b_{kj_0})$ the desirable and undesirable outputs are expanded and contracted, respectively, according to a direction that corresponds to proportional changes to the original levels for each DMU. In order to facilitate the interpretation of the DMUs' projection to the frontier, Figure 4.2 illustrates the directional vectors for DMUs A and B (gA and gB) and the DMUs' projection on the frontier, corresponding to points A* and B*, respectively.

The value of the directional distance function (corresponding to an ineffi-

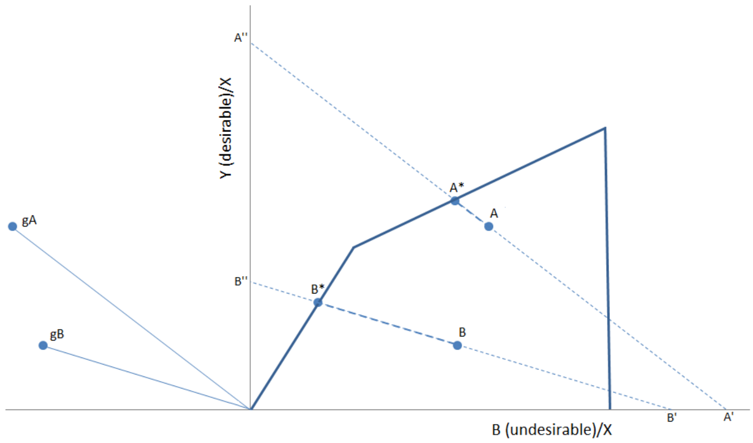


Figure 4.2: Production possibility set for the directional distance function

ciency measure) is given by β in the optimal solution of model (4.27). For DMU A it corresponds to the ratio $\frac{AA^*}{AA'}$ (taking the axis X as reference), or equivalently, by the ratio $\frac{AA^*}{AA''}$ (taking the axis Y as reference). Similarly, for DMU B it is given by the ratio $\frac{BB^*}{BB'}$, or equivalently, by the ratio $\frac{BB^*}{BB''}$.

The value of the directional distance function in assessments with undesirable outputs is always greater than or equal to zero, with values greater than zero signalling the existence of inefficiencies, and zero meaning that the production occurs on the frontier of the production possibility set.

4.3.2 The Malmquist-Luenberger index

Chung et al. (1997) defined the Malmquist-Luenberger index with the aim to define a productivity index that is able to credit a firm for reductions in undesirable outputs without requiring changes in their original measurement scale, as shown in (4.28). It is derived from the output oriented Malmquist index, shown in (4.14), using the equivalence between the Shephard's output distance function and the directional distance function, shown in (4.9).

$$ML_o^{t,t+1} = \left[\frac{(1+\vec{D}^t(x^t,y^t,b^t;y^t,-b^t))}{(1+\vec{D}^t(x^{t+1},y^{t+1},b^{t+1};y^{t+1},-b^{t+1}))}\frac{(1+\vec{D}^{t+1}(x^t,y^t,b^t;y^t,-b^t))}{(1+\vec{D}^{t+1}(x^{t+1},y^{t+1},b^{t+1};y^{t+1},-b^{t+1}))}\right]^{\frac{1}{2}} \tag{4.28}$$

The Malmquist-Luenberger index can be decomposed as follows:

$$EC_o^{t,t+1} = \frac{1+\vec{D}^t(x^t,y^t,b^t;y^t,-b^t)}{1+\vec{D}^{t+1}(x^{t+1},y^{t+1},b^{t+1};y^{t+1},-b^{t+1})} \tag{4.29}$$

and

$$TC_o^{t,t+1} = \left[\frac{(1+\vec{D}^{t+1}(x^t,y^t,b^t;y^t,-b^t))}{(1+\vec{D}^t(x^t,y^t,b^t;y^t,-b^t))}\frac{(1+\vec{D}^{t+1}(x^{t+1},y^{t+1},b^{t+1};y^{t+1},-b^{t+1}))}{(1+\vec{D}^t(x^{t+1},y^{t+1},b^{t+1};y^{t+1},-b^{t+1}))}\right]^{\frac{1}{2}} \tag{4.30}$$

The estimation of within-period and mixed-period directional distance functions used to calculate the Malmquist-Luenberger index is obtained using model (4.27), assuming a directional vector equal to $g = (y_{rj_0}, -b_{kj_0})$. While the value of the function in the within-period assessment is always greater than or equal to zero, in the mixed-period it can be smaller, equal or greater than zero. Values smaller than zero occur when the output quantities observed in one period are not feasible in the technology corresponding to the other period. Values of the Malmquist-Luenberger index greater than one correspond to productivity improvements, whereas values smaller than one signal productivity decline.

Note that the Malmquist-Luenberger index is calculated using directional distance functions specified with a vector that seeks for simultaneous improvements to both desirable and undesirable outputs. The value of the ML index (4.28) is identical to the output oriented Malmquist index when the directional vector is specified to improve desirable outputs only. In case the directional vector is specified to seek for improvements in both types of outputs, the directional distance function is not equivalent to Shephard's distance function, as the latter cannot account for simultaneous changes to desirable and undesirable outputs. Therefore, the Malmquist-Luenberger and the Malmquist indices can only be considered comparable, as noted by Chung et al. (1997).

4.3.3 The Luenberger productivity index

As an alternative to the Malmquist-Luenberger indices previously discussed, it is possible to use the Luenberger productivity index to assess productivity change over time. Chambers (1996) introduced the Luenberger productivity index as a difference of directional distance functions.

The Luenberger productivity index for an assessment involving inputs x_{ij} $(i = 1, ..., m)$, desirable outputs y_{rj} $(r = 1, ..., s)$ and undesirable outputs b_{kj} $(k = 1, ..., l)$, is defined as follows:

$$\begin{aligned} L^{t,t+1} = \tfrac{1}{2}\big[&\vec{D}^{t+1}(x^t,y^t,b^t;y^t,-b^t)-\vec{D}^{t+1}(x^{t+1},y^{t+1},b^{t+1};y^{t+1},-b^{t+1}) \\ &+\vec{D}^t(x^t,y^t,b^t;y^t,-b^t)-\vec{D}^t(x^{t+1},y^{t+1},b^{t+1};y^{t+1},-b^{t+1})\big] \end{aligned} \tag{4.31}$$

$L^{t,t+1}$ measures the productivity change between time periods t and $t+1$. Following the idea used to construct the Malmquist index, in order to avoid an arbitrary choice between the base periods, the Luenberger productivity index is given by an arithmetic mean of the indices corresponding to the periods t (the first difference) and $t+1$ (the second difference inside the square brackets).

As explained in Fare and Grosskopf (2005), the Luenberger productivity index can be addictively decomposed in two components: efficiency change and technological change, as shown in (4.32) and (4.33), respectively. This decomposition is similar to the one proposed by Fare et al. (1994b) in the context of the Malmquist index.

$$LEC^{t,t+1} = \vec{D}^{t}(x^{t},y^{t},b^{t};y,-b) - \vec{D}^{t+1}(x^{t+1},y^{t+1},b^{t+1};y^{t+1},-b^{t+1}) \tag{4.32}$$

$$\begin{aligned} LTC^{t,t+1} = \tfrac{1}{2}\big[&\vec{D}^{t+1}(x^{t+1},y^{t+1},b^{t+1};y^{t+1},-b^{t+1}) - \vec{D}^{t}(x^{t+1},y^{t+1},b^{t+1};y^{t+1},-b^{t+1}) \\ &+ \vec{D}^{t+1}(x^{t},y^{t},b^{t};y^{t},-b^{t}) - \vec{D}^{t}(x^{t},y^{t},b^{t};y^{t},-b^{t})\big] \end{aligned} \tag{4.33}$$

Both the Luenberger productivity index and its components signal improvements with values greater than zero, and declines in productivity with values smaller than zero. Values equal to zero indicate no productivity change. The change in relative efficiency between periods t and t+1 can be interpreted as the change in the distance between the observed production level and the maximum potential production. As in the Malmquist index and the ML indices, improvements in the efficiency change are evidence of catching up to the frontier, and the technological change component captures the shift in technology between the two periods of time considered.

4.4 An enhanced version of the Malmquist-Luenberger index

In this chapter, we propose an alternative formulation of the Malmquist-Luenberger index, as shown in (4.34), which is derived from the input oriented Malmquist index, shown in (4.18), using the equivalence between the distance functions shown in (4.11).

$$ML_{i}^{t,t+1} = \left[\frac{(1-\vec{D}^{t}(x^{t+1},y^{t+1},b^{t+1};y^{t+1},-b^{t+1}))}{(1-\vec{D}^{t}(x^{t},y^{t},b^{t};y^{t},-b^{t}))}\,\frac{(1-\vec{D}^{t+1}(x^{t+1},y^{t+1},b^{t+1};y^{t+1},-b^{t+1}))}{(1-\vec{D}^{t+1}(x^{t},y^{t},b^{t};y^{t},-b^{t}))}\right]^{\frac{1}{2}} \tag{4.34}$$

The components of the Malmquist-Luenberger index, shown in (4.34), are as follows:

$$EC_i^{t,t+1} = \frac{1-\vec{D}^{t+1}(x^{t+1},y^{t+1},b^{t+1};y^{t+1},-b^{t+1})}{1-\vec{D}^{t}(x^{t},y^{t},b^{t};y^{t},-b^{t})} \tag{4.35}$$

and

$$TC_i^{t,t+1} = \left[\frac{(1-\vec{D}^{t}(x^{t},y^{t},b^{t};y^{t},-b^{t}))}{(1-\vec{D}^{t+1}(x^{t},y^{t},b^{t};y^{t},-b^{t}))}\frac{(1-\vec{D}^{t}(x^{t+1},y^{t+1},b^{t+1};y^{t+1},-b^{t+1}))}{(1-\vec{D}^{t+1}(x^{t+1},y^{t+1},b^{t+1};y^{t+1},-b^{t+1}))}\right]^{\frac{1}{2}} \tag{4.36}$$

This alternative version of the Malmquist-Luenberger index ($ML_i^{t,t+1}$), which is input-oriented, is obtained using a directional distance function specified with a vector equal to $g = (y_{rj_0}, -b_{kj_0})$. This index would only be equivalent to the input oriented Malmquist index if the directional vector was specified in order to reduce only the undesirable outputs, keeping the desirable outputs with their current value.

Note that to preserve the interpretation of the productivity indices such that values greater than one mean improvements, the input oriented versions of the Malmquist index and Malmquist-Luenberger index use the inverse of the input distance function, corresponding to an efficiency measure. For the output oriented indices, the output distance function is equal to the efficiency measure.

The Malmquist-Luenberger indices ($ML_i^{t,t+1}$ and $ML_o^{t,t+1}$) are derived from the Malmquist indices ($MI_i^{t,t+1}$ and $MI_o^{t,t+1}$) based on the relations between the directional distance function and the Shephard's distance functions, presented in (4.9) and (4.11). The estimation of the Shephard's distance functions can also be visualized in Figure 4.2, which was previously used in section 4.3.1 to illustrate the directional distance function. The value of the Shephard's output distance function (efficiency measure) used to calculate the $ML_o^{t,t+1}$ index for the DMU A, is given by the ratio $\frac{AA'}{A^*A'}$ (corresponding to the relation $D_o(x^t, y^t) = 1/(1 + \vec{D}^t(x^t, y^t, b^t; y^t, -b^t))$. On the other hand, the inverse of the Shephard's input distance function (efficiency measure) used to calculate the $ML_i^{t,t+1}$ index for the DMU A, is given by $\frac{A^*A''}{AA''}$ (corresponding to the relation $(D_i^t(x^t, y^t))^{-1} = 1 - \vec{D}^t(x^t, y^t, b^t; y^t, -b^t))$.

Although both ratios are acceptable to estimate an efficiency measure for DMU A (as they only differ in the axis used as reference to estimate efficiency), they do not result in the same values. As a consequence, the results of the input-based and output-based Malmquist-Luenberger indices are different, as they depend on the relation used to convert the directional distance function in a Shephard's distance function. Recall that, as noted by Chung et al. (1997), the Shephard's distance functions and the directional distance functions are not equivalent in the presence of both desirable and undesirable outputs, so this implies that the input-based and output-based ML indices are not identical.

Following the idea of Fare et al. (1994b), that in order to avoid the use of an arbitrary reference technology proposed the estimation of the Malmquist index as the geometric mean of two Malmquist indices using as reference the technology in different time periods (t and $t+1$), we propose an enhanced version of the Malmquist-Luenberger index that avoids an arbitrary choice between the input or output Shephard's distance function used to derive the Malmquist-Luenberger index. The new index is specified as the geometric mean of the indices $ML_o^{t,t+1}$ and $ML_i^{t,t+1}$, as shown in (4.37).

$$ML_{av}^{t,t+1} = \left[ML_o^{t,t+1} \times ML_i^{t,t+1}\right]^{\frac{1}{2}} \tag{4.37}$$

The new Malmquist-Luenberger index can be decomposed as follows:

$$EC_{av}^{t,t+1} = \left[EC_o^{t,t+1} \times EC_i^{t,t+1}\right]^{\frac{1}{2}} \tag{4.38}$$

and

$$TC_{av}^{t,t+1} = \left[TC_o^{t,t+1} \times TC_i^{t,t+1}\right]^{\frac{1}{2}} \tag{4.39}$$

The three versions of the ML indices (4.28), (4.34) and (4.37), corresponding to the input, output or average indices, indicate improvement, stagnation and decline in productivity, by values greater, equal or smaller than one, respectively.

Next we illustrative graphically how the productivity indices previously discussed (the three versions of the Malmquist-Luenberger index and the Luenberger index) measure changes in productivity over time in the presence of undesirable outputs. The advantages and limitations of each approached are also discussed.

4.5 Empirical examples

4.5.1 Graphical illustration of the productivity change

Our illustrative example is based on a set of 6 DMUs assessed in two time periods: 0 and 1. To allow a graphical illustration of the models, these DMUs are assessed considering two output indicators: Y, a desirable output, and B, an undesirable output, and a unitary input underlying the evaluation of every DMU in both time periods. Table 4.1 shows the data for the 6 DMUs.

Table 4.1: Data for the illustrative example

DMUs	X^0	Y^0 (desirable)	B^0 (undesirable)	X^1	Y^1 (desirable)	B^1 (undesirable)
A	1	8	12	1	12	9
B	1	22	30	1	25	34
C	1	18	10	1	30	13
D	1	7	20	1	10	21
E	1	30	34	1	35	29
F	1	20	23	1	24	24

Figure 4.3 illustrates the production possibility sets for the illustrative example, in time periods 0 and 1, corresponding to an evaluation using the

directional distance function model (4.27). The projections to the frontier for DMUs D and F, in time period 0 and 1, are also represented. Note that the DMUs are projected to the frontier according to a direction that corresponds to proportional changes to the original output values of each DMU. As lower values in the output indicator B correspond to better performance, the technically efficient frontier is given by the segments linking the origin and the DMUs C^0 and E^0 for the first time period, and the origin and the DMUs C^1 and E^1 for the second time period.

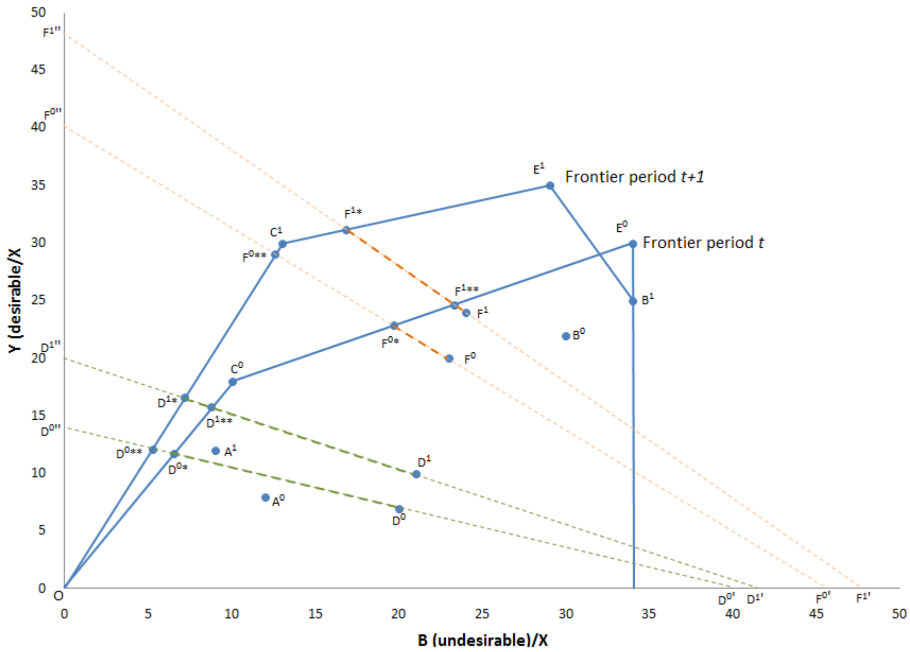


Figure 4.3: Production possibility set for the illustrative example in time periods 0 and 1

The values of the directional distance functions (that corresponds to the inefficiency measure β in model (4.27)) used to calculate the Malmquist-Luenberger indices and the Luenberger index are shown in Table 4.2. For DMU F, in time period 0, the value of the directional distance function $\vec{D}^0(x^0, y^0, b^0; y^0, -b^0)$ is 0.1429 and $\vec{D}^1(x^0, y^0, b^0; y^0, -b^0)$ is 0.4526, corresponding to the proportional improvement required to the outputs of this

DMU in order to operate on the frontier of the production possibility set in periods 0 and 1, respectively (i.e. to reach the points F^{0*} and F^{0**}, respectively, in Figure 4.3).

Table 4.2: Value of the Directional Distance Function for the illustrative example

DMUs	$\vec{D}^0(x^0, y^0, b^0; y^0, -b^0)$	$\vec{D}^0(x^1, y^1, b^1; y^1, -b^1)$	$\vec{D}^1(x^1, y^1, b^1; y^1, -b^1)$	$\vec{D}^1(x^0, y^0, b^0; y^0, -b^0)$
A	0.4595	0.1489	0.2676	0.5517
B	0.1622	0.1190	0.3246	0.4243
C	0	-0.2877	0	0.1236
D	0.6744	0.5816	0.6579	0.7366
E	0	-0.1515	0	0.1615
F	0.1429	0.0278	0.2996	0.4526

Considering the inefficiency score β for DMU F in period 0, equal to 0.1429, and using the axis X as reference, the value of the directional distance function corresponds to the ratio $\frac{F^0F^{0*}}{F^0F^{0'}}$. On the other hand, taking as reference the axis Y, the directional distance function can alternatively be obtained by the ratio $\frac{F^0F^{0*}}{F^0F^{0''}}$. Both representations give the same value of the inefficiency score β (or directional distance function).

As discussed in the previous sections, a directional distance function specified using a vector that seeks for simultaneous improvements in desirable and undesirable outputs can be related to both the input and output Shephard's distance functions. This means that, in Figure 4.3, the directional distance function can be converted to a distance function using a projection in relation to axis X (for an output orientation) or axis Y (for an input orientation). Using the axis X as reference, the efficiency for the DMU F is given by the ratio $\frac{F^0F^{0'}}{F^{0*}F^{0'}} = 0.8750$, that corresponds to the relationship $D_o^0(x^0, y^0) = 1/(1 + \vec{D}^0(x^0, y^0, b^0; y^0, -b^0))$ used to define the index $ML_o^{t,t+1}$. On the other hand, when the axis Y is used as reference, the efficiency measure of DMU F is given by $\frac{F^{0*}F^{0''}}{F^0F^{0''}} = 0.8571$, corresponding to the relationship $(D_i^0(x^0, y^0))^{-1} = 1 - \vec{D}^0(x^0, y^0, b^0; y^0, -b^0)$ used to define the index $ML_i^{t,t+1}$.

As in ML indices the directional distance function is specified using a vector that seeks for increases in desirable outputs and reductions in undesirable outputs, both indices $ML_o^{t,t+1}$ and $ML_i^{t,t+1}$ can be considered equally good to estimate changes in productivity over time. Therefore, in order to avoid an arbitrary selection between the indices $ML_o^{t,t+1}$ and $ML_i^{t,t+1}$, their geometric mean can be used to calculate a new ML index, as proposed in expression (4.37). The results for the three different expressions of the ML index and for the Luenberger index are presented in Table 4.3.

Table 4.3: Results for the different expressions of the productivity indices

DMUs	ML_o (rank)	EC	TE	ML_i (rank)	EC	TE	ML_{av} (rank)	EC	TE	L (rank)	LEC	LTC
A	1.247 (5)	1.151	1.083	1.604 (6)	1.355	1.184	1.414 (6)	1.249	1.132	0.297 (6)	0.192	0.105
B	1.057 (2)	0.877	1.204	1.111 (1)	0.806	1.378	1.083 (1)	0.841	1.288	0.071 (1)	-0.162	0.234
C	1.256 (6)	1	1.256	1.212 (4)	1	1.212	1.234 (5)	1	1.234	0.206 (5)	0	0.206
D	1.053 (1)	1.010	1.043	1.292 (5)	1.051	1.230	1.166 (3)	1.030	1.132	0.086 (2)	0.016	0.069
E	1.170 (4)	1	1.170	1.172 (2)	1	1.172	1.171 (4)	1	1.171	0.156 (4)	0	0.156
F	1.115 (3)	0.879	1.268	1.205 (3)	0.817	1.474	1.159 (2)	0.848	1.367	0.134 (3)	-0.157	0.291

Comparing the results obtained for the different formulations of the ML indices and their components, we can see that not only the magnitude of the productivity change estimate is different but also the ranking of DMUs varies. Note that the ranking of DMUs given by the Average ML index is very similar to the ranking of DMUs given by the Luenberger productivity index. This suggests that the Average ML index is more aligned with the Luenberger index than the other two versions of the ML index ($ML_o^{t,t+1}$ and $ML_i^{t,t+1}$).

In the next section, the productivity indices discussed in this chapter are applied to a real-world assessment, which allows comparing the Malmquist-Luenberger indices with the Luenberger index, and verify the robustness of the conclusions that can be drawn using different measures.

4.5.2 European commercial transport industry

Our illustrative application uses panel data describing the commercial transport industry of 17 countries, 16 members of the European Union and Norway, explored in the paper of Krautzberger and Wetzel (2012). The authors used the ML index proposed by Chung et al. (1997) to compare the CO_2-sensitive productivity development of the European commercial transport industry between 1995 and 2006. Krautzberger and Wetzel (2012) conducted the assessment considering as input variables *capital stock*, *number of employees* and *intermediate inputs* accounting for expenses with energy, materials and services, with one desirable output (*gross output*) and one undesirable output (CO_2 *emissions*). For more details concerning the selection of the inputs and outputs and data sources, see Krautzberger and Wetzel (2012).

As this assessment includes an undesirable output, it is appropriate to conduct the comparison between the three different ML indices (expressions (4.28), (4.34) and (4.37)), and the Luenberger Index (expression (4.31)). In our application, we selected only the last two years (2005 and 2006) to illustrate the productivity indices. The data used in this empirical application is shown in Table A.1 in Appendix A.

Table 4.4 shows the results of the productivity change indices, and their components of efficiency change and technological change, obtained for the 17 countries analysed. Denmark was the only country that yielded an infeasible solution in the mixed-period assessments. In assessments involving undesirable outputs, using either ML indices or the Luenberger index, infeasibilities occur in cases in which a DMU from one period is beyond the production possibility set of the other time period and its projection is in a direction where the frontier of the other time period does not exist. In this case it is not possible to provide an estimate of productivity change. Further details on this issue can be found in Aparicio et al. (2013). Note that this problem only occurs in the Malmquist index when the distance function in estimated assuming variable returns to scale.

Table 4.4: Countries' productivity change for the three versions of the ML index and the Luenberger index

Countries	ML_o (rank)	EC	TE	ML_i (rank)	EC	TE	ML_{av} (rank)	EC	TE	L (rank)	LEC	LTC
AT	1.031 (13)	1.019	1.012	1.042 (14)	1.026	1.016	1.037 (14)	1.023	1.014	0.035 (13)	0.022	0.013
BE	1.031 (12)	1	1.031	1.030 (12)	1	1.030	1.031 (12)	1	1.031	0.030 (12)	0	0.030
CZ	1.043 (16)	1.036	1.006	1.045 (15)	1.038	1.006	1.044 (16)	1.037	1.006	0.043 (16)	0.036	0.006
DE	1.024 (11)	1.025	0.999	1.026 (11)	1.027	0.999	1.025 (11)	1.026	0.999	0.025 (11)	0.025	-0.001
DK	n/a n/a	1	n/a	n/a n/a	1	n/a	n/a n/a	1	n/a	n/a n/a	0	n/a
EE	0.983 (4)	1	0.983	0.985 (5)	1	0.985	0.984 (5)	1	0.984	-0.016 (5)	0	-0.016
ES	1.010 (9)	0.999	1.011	1.013 (9)	0.999	1.014	1.012 (9)	0.999	1.012	0.011 (9)	-0.001	0.012
FI	0.985 (5)	1	0.985	0.986 (6)	1	0.986	0.985 (6)	1	0.985	-0.015 (6)	0	-0.015
FR	0.996 (7)	1	0.996	0.996 (7)	1	0.996	0.996 (7)	1	0.996	-0.004 (7)	0	-0.004
HU	0.985 (6)	1.015	0.971	0.983 (4)	1.017	0.966	0.984 (4)	1.016	0.968	-0.016 (4)	0.016	-0.032
IT	1.017 (10)	1	1.017	1.016 (10)	1	1.016	1.017 (10)	1	1.017	0.017 (10)	0	0.017
NL	1.034 (14)	1.013	1.021	1.051 (16)	1.019	1.031	1.043 (15)	1.016	1.026	0.040 (15)	0.015	0.025
NO	1.003 (8)	0.992	1.012	1.004 (8)	0.99	1.015	1.004 (8)	0.991	1.013	0.004 (8)	-0.009	0.013
SE	1.037 (15)	1	1.037	1.035 (13)	1	1.035	1.036 (13)	1	1.036	0.036 (14)	0	0.036
Sl	0.958 (1)	0.968	0.99	0.952 (1)	0.963	0.988	0.955 (1)	0.966	0.989	-0.045 (1)	-0.035	-0.011
SK	0.979 (3)	0.972	1.008	0.978 (3)	0.971	1.008	0.979 (3)	0.971	1.008	-0.022 (3)	-0.029	0.008
UK	0.963 (2)	1	0.963	0.966 (2)	1	0.966	0.964 (2)	1	0.964	-0.036 (2)	0	-0.036

Austria (AT), Belgium (BE), Czech Republic (CZ), Germany (DE), Denmark (DK),Estonia (EE), Spain (ES), Finland (FI), France (FR), Hungary (HU), Italy (IT), Netherlands (NL), Norway (NO), Sweden (SE), Slovenia (SI), Slovakia (SK), and United Kingdom (UK).

The results for the different indices ($ML_o^{t,t+1}$, $ML_i^{t,t+1}$, $ML_{av}^{t,t+1}$ and $L^{t,t+1}$) and their components are consistent in signalling improvements, stagnation or declines in productivity for all countries analysed. Among the 16 countries considered, 9 countries signalled improvements in productivity (ML indices greater than one and Luenberger index greater than zero) and 7 signalled declines (ML indices smaller than one and Luenberger index smaller than zero).

The Pearson correlation coefficients revealed a strong correlation between the values of the three versions of the Malmquist-Luenberger indices ($ML_o^{t,t+1}$, $ML_i^{t,t+1}$ and $ML_{av}^{t,t+1}$) and the Luenberger index. The value of the correlation coefficients are equal to 0.9975, 0.9961 and 0.9998, respectively.

The number of countries that did not obtain the same ranking as the Luenberger index was 5 for the $ML_o^{t,t+1}$ index, and 4 for the $ML_i^{t,t+1}$ index. The Average ML index only had 2 countries with a different rank from that of the Luenberger index. The Kendall rank-order correlation coefficients between the ranking of countries given by the three version of the Malmquist-Luenberger index ($ML_o^{t,t+1}$, $ML_i^{t,t+1}$ and $ML_{av}^{t,t+1}$) and the Luenberger index are equal to 0.9500, 0.9667 and 0.9833, respectively. The Kendall coefficient is a nonparametric measure of correlation between two ranked variables. It measures the difference between the probability that the observed data are in the same order and the probability that the observed data are not in the same order.

Although all coefficients indicate strong correlation between the indices (all of them are significantly different from zero with a p-value $\leq$ 0.0000), the correlation coefficients (Pearson and Kendall) between the Average ML index and the Luenberger index are almost equal to one, suggesting that the new Average ML index is more aligned with the Luenberger index than the original version of the ML index, based on the output distance function.

The correlation between the indices is illustrated in Figure 4.4. The plot on the left presents the correlation between the values of the indices, and the plot on the right presents the correlation between the ranking of countries according to their productivity change. In general, the dispersion is higher among countries that presented the largest increases in productivity.

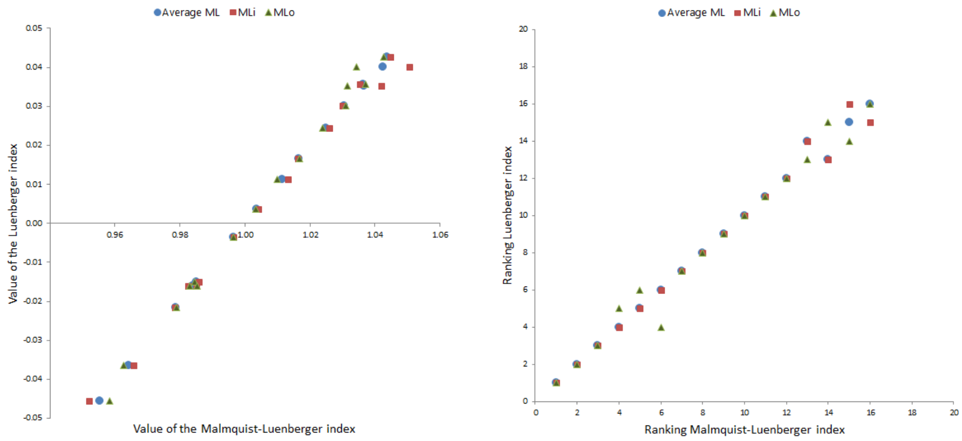


Figure 4.4: Correlation between the ML indices and the Luenberger index

Boussemart et al. (2003) and Managi (2003) studied the relationship between the Malmquist index and the Luenberger index. The authors showed that the logarithm of the estimate of productivity change obtained using the Malmquist index has a magnitude that is nearly twice the value of the Luenberger index. In our empirical example, we found that the logarithm of the Average ML productivity index is approximately equal to the Luenberger productivity index ($ln(ML_{av}^{t,t+1}) \approx L^{t,t+1}$). For example, France obtained a score of -0.00357 in the Luenberger productivity index, and this value is equal to the logarithm of the Average Malmquist-Luenberger productivity index ($\ln(0.99644)=-0.00357$) considering a precision with 3 significant digits. Considering this relationship, the sum of the squared errors, for all countries, between the logarithm of the three versions of the ML indices, $ML_o^{t,t+1}$, $ML_i^{t,t+1}$ and $ML_{av}^{t,t+1}$, and the Luenberger productivity index are equal to 0.000080, 0.000146 and 0.000003, respectively. As the values of the changes in productivity between the two periods decrease, the relationship becomes more accurate.

4.6 Conclusions

This study addressed the different approaches that have been used to assess productivity change in the presence of undesirable outputs: the ratio-based Malmquist-Luenberger index and the difference-based Luenberger index. It was demonstrated that the ML index can be derived from a standard Malmquist index using the relationship between the directional distance function and the Shephard's output distance function, or equivalently, using the relationship between the directional distance function with the Shephard's input distance function. The two versions of the ML indices represent equally good adaptations of the Malmquist index for assessments involving both desirable and undesirable outputs. In order to avoid the need to arbitrarily choose one of the measures, it was proposed a new version of the ML index, given by the geometric mean of these alternative versions of the ML indices.

Both the Malmquist-Luenberger and Luenberger indices are estimated using directional distance functions, and thus they can accommodate undesirable outputs without requiring changes in their original measurement scale. The main methodological difference between the Malmquist-Luenberger and Luenberger indices, which can motivate the use of one or another, is related to their multiplicative or additive nature.

In cases in which the researcher has a preference for ratio-based indices, and the assessment involves simultaneous improvements to inputs and desirable and undesirable outputs, we suggest using the new version of the Malmquist-Luenberger index proposed in this study, as it has the advantage of incorporating both orientations in the computation of the productivity change score, and thus represent more accurately the changes in DMUs features.

Using an empirical example, we compared the results obtained by the different versions of the ML indices with the Luenberger index, which is an established difference-based measure to assess productivity change over time considering simultaneous adjustments to inputs and desirable and undesirable outputs. The empirical results suggested that the Average ML index is more aligned with the Luenberger index than the other two versions of the ML index, both in terms of the relationship between the values of the productivity estimates as well as in terms of the rankings obtained. The results also showed that the logarithm of the Average Malmquist-Luenberger index is approximately equal to the Luenberger index.

Although the ML and the Luenberger indices discussed in this chapter were originally proposed for assessments involving the conversion of inputs to desirable and undesirable outputs, they can be easily adapted for the context of composite indicators. By assuming a unitary level of inputs for all DMUs in the model (4.27), used to calculate both the ML and the Luenberger indices, it becomes the Direction CI model presented in (3.6), as explained in section 3.3.2.

Chapter 5

Benchmarking countries' environmental performance

5.1 Introduction

Environmental concerns have increased dramatically in the past few years and are now among the most serious challenges affecting people's wellbeing. Besides the climate change that has been widely discussed, other environmental problems such as local air and water pollution, soil erosion, water scarcity, deforestation, and loss of biodiversity are also becoming more serious (World Bank, 2008).

Countries are facing new challenges to control their waste production and to reduce the consumption of natural resources, in order to achieve the environmental targets imposed by international agreements such as the Kyoto Protocol (United Nations, 1998) or the European Union climate and energy package (European Union, 2008). In this context, it is imperative that countries become able to monitor their environmental performance in order to understand how they are doing compared to others and to identify the potential for improvements.

Environmental performance assessments are often conducted using environmental indicators that are able to measure the pressures on the environment, to appraise the state of the ecosystem and to evaluate the impacts on human activity resulting from changes in environmental quality. These indicators usually measure particular features of the environment and provide a starting point for performance assessments. It is also common to use composite indicators (CI) to aggregate several individual indicators in a summary measure of performance. Although the indicators, individual or aggregated, can be extremely useful to guide discussions on environmental issues and attract public interest, they do not provide guidelines that countries should follow to improve performance. This chapter aims to assess countries environmental performance using an enhanced CI model that, besides assigning a summary measure of performance for each country, can be used for benchmarking purposes. The CI is defined based on the Data Envelopment Analysis (DEA) technique.

In the context of an environmental performance assessment, the DEA models can be used either to measure the environmental efficiency of a given set of units, i.e. the ability of convert inputs to outputs, or to provide an environmental effectiveness measure, which aggregates several output indicators in a CI (i.e. looking only at the achievements, rather than the conversion of inputs to outputs). Irrespectively of the approach followed in the environmental performance assessment, both desirable and undesirable outputs may be present. Since the standard DEA models rely on the assumption that outputs are maximized, the undesirable outputs must be dealt with in order to be accommodated in a DEA formulation.

As discussed in chapter 3, two different approaches can be followed to treat undesirable outputs, a direct and an indirect approach. This chapter illustrate the application of the indirect approach, which is based on a transformation of the measurement scale of the undesirable outputs. Moreover, it is illustrated the use of the assurance region type I weight restrictions, described in section (3.4.2), to incorporate in the CI model information of the relative importance of indicators, expressed as a percentage.

5.2 Review of composite indicators for environmental performance assessment

In the past few years, efforts to assess environmental performance of organizations, cities and countries have generated a large number of indicators related to gas emissions, water quality, green space area, waste production, among others. Due to the large amount of individual indicators available, the process of analyzing and understanding the information available becomes difficult. Therefore, many of these indicators are often aggregated into composite indicators to gain an overall picture of performance that can be used by decision makers for planning and control purposes.

According to the Organisation for Economic Co-operation and Development report by Nardo et al. (2008), the construction of composite indicators involves several stages: the selection of sub-indicators, the treatment of missing values, the understanding of the data, and the specification of the weights for the sub-indicators to be used in the aggregation model.

Concerning the selection of sub-indicators, they should be selected according to their analytical soundness, measurability, coverage, and relevance to the phenomenon being assessed. Regarding to the treatment of missing values, there are two possibilities: to impute values to replace the missing fields or to remove the whole observation with missing data from the analysis. Concerning the understanding of the data, an exploratory analysis should be performed to study the overall structure of the dataset. It is necessary to understand the relationship among indicators and observations, and to identify outliers and extreme values that can introduce bias in the results.

Concerning the specification of weights for the sub-indicators, they can be specified based on quantitative methods or expert judgment. The weights can reflect policy priorities, or reward the factors deemed to be more important to the performance assessment. The choice of weights inevitably impacts the results of the CI, so the method should be robust to avoid undermining the credibility of the CI results. Since the specification of the weights is often subject of criticism and disagreement, most composite indicators rely on equal weighting to minimize the subjectivity. One such example is the calculation of the Human Development Index (United Nations Development Program, 2000), which is based on the use of equal weights for its component indices, which reflect longevity (measured by life expectancy at birth), education attainment (measured by adult literacy and enrolment rate) and standard of living (measured by GDP per capita).

In the context of environmental performance assessments, the Climate Change Performance Index (CCPI), and the Environmental Performance Index (EPI) are examples of well-established composite indicators that aggregate individual indicators in a summary measure.

The CCPI is measured via 12 different indicators, which can be classified in the following categories: emissions trend, emissions level and climate policy. This index compares countries that together are responsible for more than 90% of global energy-related CO_2 emissions. The CCPI needs the predefinition of weights for the indicators. The countries ranking is calculated from the weighted average of the scores achieved by the countries evaluated in the indicators considered (Burck et al., 2009).

The EPI provides a global index based in 25 indicators, grouped in 10 categories, covering two core objectives: Environmental Health and Ecosystem Vitality. The first core objective measures the environmental effects on human health, whereas the second measures the state of the ecosystem and the natural resources management. Each of these two core objectives contributes with a weight of 50% to the overall EPI score. The EPI construction requires a predefinition of weights and targets for the indicators. The weights assigned to the indicators are determined through expert judgment. For each country and each indicator, a proximity-to-target value is calculated based on the gap between the current results presented by each country and a target previously identified. The targets are defined based on four sources: treaties or other internationally agreed goals, standards defined by international organizations, national regulatory requirements, or expert judgment. After defining the indicators' weights and the proximity-to-target of each country, a weighted average is calculated to obtain the EPI score for each country (Emerson et al., 2010). The countries' performance assessment presented in our study was conducted using the indicators that underlie the estimation of the EPI 2010. The EPI indicators and weights are reported on Table 5.1. For further details on the specification of the EPI indicators see the metadata information in Emerson et al. (2010).

Both the EPI and CCPI use a weighted average to provide an overall measure of performance, and rely on expert opinion to specify the weights. An alternative to overcome the difficulties in the selection of weights is to use Data Envelopment Analysis (DEA) to determine the weights. Using DEA, individual indicator weights result from an optimizing process, based on linear programming, so they are less prone to subjectivity and controversy.

Although the DEA technique has been used to construct CI in different fields, such as the evaluation of urban quality of life (Morais and Camanho, 2011), human development (Mahlberg and Obersteiner, 2001; Despotis, 2004, 2005), social deprivation (Zaim et al., 2001) , technology achievement (Cherchye et al., 2008), monetary aggregation (Sahoo and Acharya, 2010) or the financial soundness of construction companies (Horta et al., 2010), its application for environmental performance assessment is scarce. This chapter

Table 5.1: EPI indicators

Policy Categories	Indicators	Weight (w)
Environmental burden of disease	Disability Life Adjusted Years	25%
Water (effects on humans)	Access to adequate sanitation	6.3%
	Access to drinking water	6.3%
Air pollution (effects on humans)	Indoor air pollution	6.3%
	Outdoor air pollution - Urban Particulates	6.3%
Air Pollution (effects on ecosystem)	Ozone Exceedance	0.7%
	Non-methane volatile organic compound emissions	0.7%
	Sulfur dioxide emissions	2.1%
	Nitrogen oxides emissions	0.7%
Water (effects on ecosystem)	Water quality index	2.1%
	Water stress index	1.0%
	Water scarcity index	1.0%
Biodiversity e Habitat	Biome protection	2.1%
	Critical habitat protection	1.0%
	Marine protection	1.0%
Forestry	Growing stock change	2.1%
	Forest cover change	2.1%
Fisheries	Marine trophic index	2.1%
	Trawling intensity	2.1%
Agriculture	Agricultural water intensity	0.8%
	Agricultural subsidies	1.3%
	Pesticide regulation	2.1%
Climate Change	Greenhouse gas emissions per capita	12.5%
	Industrial greenhouse gas emissions intensity	6.3%
	CO_2 emissions per electricity generation	6.3%

contributes to the literature in this field by proposing a methodology to evaluate countries environmental performance based on the construction of a composite indicator whose weights are specified using DEA.

5.3 Methodology

Three main features of the DEA technique motivated its use in this chapter to define a CI to assess countries environmental performance. The first is related to the ability to identify best-practice peers corresponding to countries with a similar profile to that of the country under assessment. The second is related to the possibility of assigning weights to individual indicators recurring to optimization. This procedure has the advantage of being less prone to subjectivity, and allowing the identification of the areas in which countries have good or bad performance. Finally, DEA is able to handle data measured in different measurement scales and thus it is possible to use the raw data corresponding to each EPI indicator, without prior normalisations or conversions to similar measurement scales.

In the environmental performance assessment presented in this chapter, the EPI indicators shown in Table 5.1 were included as outputs of the CI model (3.2), developed in chapter 3. As all outputs were measured as ratios or indexes, we considered an identical input level for all DMUs.

In a first moment, in order to establish a ranking of countries environmental performance, we fixed the weights of all indicators in the CI model to ensure that all countries are evaluated using the same criteria (i.e., using common weights). We adopted weight restrictions that mimic the value judgments implicit in the EPI, which are based on expert opinion (see w in Table 5.1). The use of common weights allows obtaining a robust ranking of countries, as it prevents obtaining a high performance score only due to a judicious choice of weights. Furthermore, the use of a unique weighting system for all DMUs improves the discrimination power of the performance assessment.

The weight restrictions imposed to the CI model (3.2), leading to the use of a common weighting system, are shown in (5.1).

$$\begin{cases} \dfrac{u_r \ \bar{y}_r}{\sum_{r'=1}^{s} u_{r'} \ \bar{y}_{r'} + \sum_{k'=1}^{l} p_{k'} \ (M_k - \bar{b}_{k'})} = w_r & r = 1, ..., s \\ \dfrac{p_k \ (M_k - \bar{b}_k)}{\sum_{r'=1}^{s} u_{r'} \ \bar{y}_{r'} + \sum_{k'=1}^{l} p_{k'} \ (M_k - \bar{b}_{k'})} = w_k & k = 1, ..., l \end{cases} \tag{5.1}$$

The rational underlying the specification of these restrictions is as follows. We consider an artificial DMU whose outputs are equal to the average value of each output variable (i.e., EPI indicator) in the sample. For the artificial DMU, the virtual weight of a desirable output (r) or undesirable output (k)

divided by the total virtual weight (i.e., the sum of virtual weights for all its outputs) must be equal to the EPI weight (w). As mentioned in section 3.4.2, these weight restrictions work as assurance regions type I (ARI), since they specify ratios between output weights that are not DMU-specific.

In a second moment, in order to identify the strengths and weaknesses of each country we relaxed the fixed system of weights imposed in (5.1), allowing a range of flexibility around the EPI weights (w), as shown in (5.2). In restrictions (5.2), a value of c equal to zero is the most restrictive case, leading to restrictions identical to those in (5.1). Thus, the imposition of common weights can be seen as a particular case of the more general formulation of the weight restrictions (5.2). Conversely, as the value of c increases, the constraints are relaxed, so the objective function value of the CI model will converge to an unrestricted efficiency score.

$$\begin{cases} w_r\ (1-c) \leq \dfrac{u_r\ \bar{y}_r}{\sum_{r'=1}^{s} u_{r'}\ \bar{y}_{r'} + \sum_{k'=1}^{l} p_{k'}\ (M_k - \bar{b}_{k'})} \leq w_r\ (1+c) & r = 1,...,s \\ w_k\ (1-c) \leq \dfrac{p_k\ (M_k - \bar{b}_k)}{\sum_{r'=1}^{s} u_{r'}\ \bar{y}_{r'} + \sum_{k'=1}^{l} p_{k'}\ (M_k - \bar{b}_{k'})} \leq w_k\ (1+c) & k = 1,...,l \end{cases} \tag{5.2}$$

The flexibility in the selection of the weights allows each country to show itself in a favorable light. Thus, based on the weights selected by each country, we are able to identify the areas in which countries are specialized and have better environmental performance (corresponding to the assignment of higher weights by the optimisation process), as well as the areas in which countries need to improve their performance.

The identification of peers is an important by-product of the DEA efficiency assessment. The peers can be identified through the dual formulation of the DEA model. The dual of model (3.2) complemented with the restrictions shown in (5.2), is shown in the Appendix B.

The CI model shown in (3.2), combined with the weight restrictions shown in (5.1) or (5.2), ensures that all indicators contribute to the overall performance measure, i.e. the weights of all indicators are positive. When all weights are positive, the slack variables assume a value equal to zero, as explained by Portela and Thanassoulis (2006). Therefore, the slacks do not need to be taken into account in the performance assessment based on our model.

5.4 Results and discussion

5.4.1 Exploratory data analysis

As previously explained, this study uses the indicators that underlie the estimation of the EPI to assess countries environmental performance. According to Emerson et al. (2010), the indicators used in the EPI reflect state-of-the-art data and the best current thinking in environmental health and ecological science.

The first stage of the analysis consisted on an exploratory analysis of the EPI indicators using descriptive statistics and cluster analysis. As some indicators presented extreme values that could bias the performance evaluation, these extreme values were replaced by three standard deviations from the mean.

An important issue that needs to be addressed before conducting a benchmarking exercise is related to the homogeneity of the DMUs. As argued by Dyson et al. (2001), a common pitfall in the application of DEA arises from simply attempting to compare non-homogeneous DMUs, which can be avoided by clustering the DMUs in homogeneous sets prior to the assessment. This was the approach we followed, such that the DEA model was applied to evaluate performance within clusters, ensuring comparisons only between countries with similar environmental features. As a result of this approach, the performance score of a country is obtained from a comparison only with other countries that belong to the same group.

Therefore, in the first stage of the analysis, we used cluster analysis to group the countries in relative homogeneous groups, and in the second stage we used Data Envelopment Analysis to assess the environmental performance of countries within clusters.

The cluster analysis was conducted using the k-means algorithm. The k-means is a partitionary algorithm that can be easily implemented and is able to deal with large data sets due the low running time. The algorithm finds the centroid of each cluster and assigns each object to its nearest centroid. After allocating all the objects, the values of the centroids are recalculated for each new group formed. The procedure is repeated until all the objects are well allocated in their groups, without requiring a new iteration. In other words, the process iterates until the criterion function converges. The squared error criterion (distance from the objects to their cluster centroid) is the function usually used. The k-means algorithm requires the definition of the k initial seeds (initial objects that are defined as centroid) in the first iteration. The clustering result is highly dependent of these initial seeds. A procedure often used to overcome this problem, which was also followed in our study, is to try different seeds and select the ones that produce the best value of the criterion (i.e., the lowest squared error).

As the k-means algorithm requires the prior definition of the number of clusters, we used the Davies-Bouldin index to identify the best number of clusters to split the countries. This index was proposed by Davies and Bouldin (1979) and is given by the ratio of the sum of within-cluster scatter to between-cluster separation. Partitions with smaller values for the Davies-Bouldin index are preferable. The Davies-Bouldin index applied to our sample suggested to split the countries in 4 or 7 clusters. Based on the analysis of the clusters obtained with the split in 7 groups, we believe this would be the most appropriate option, as it lead to four large clusters (with about 40 countries each), and three smaller clusters (with no more than 5 countries each). This ensured that countries with atypical behavior would not be included in the larger clusters, which enhances their homogeneity. Only the four larger clusters were analyzed in the empirical part of this chapter, as the sample size of the smaller clusters would be inappropriate for a DEA assessment.

The cluster analysis was followed by the construction of a decision tree to reveal the main features of the countries belonging to the same group. Table 5.2 shows the main characteristics found in each cluster, as well as the percentage of countries that effectively presented these characteristics.

In order to illustrate the insights that can be obtained using this procedure consider, for example, the second cluster, which mainly includes countries from Occidental Europe. The countries of this cluster are characterized by low levels of *Indoor air pollution*, *Water scarcity* and *Ozone Exceedance*, and high values for *Critical habitat protection* and *Agricultural Subsidies*. From the environmental point of view, the last variable is a bad characteristic, because the magnitude of the subsidies is positively correlated with the pressure that they exert. The first four indicators present good characteristics of the countries from this cluster. The same interpretation was done for the others clusters. The bad characteristics for each cluster are distinctly signalled in Table 5.2 with an asterisk.

5.4.2 Performance assessment results

The performance assessment was conducted using model (3.2) with a value of M_k equal to the largest value observed for the undesirable output indicator

Table 5.2: Characteristics of each cluster

Cluster	Characteristics	% of countries
C1 (36 countries): Majority of Africa, three countries of South-eastern Asia, Papua New Guinea, Nicaragua and Haiti	High Indoor air pollution* High Disability Life Adjusted Years*	94.4%
C2 (43 countries): Countries of Occidental Europe, six countries of Latin America, Russia, Japan, South Korea, Malaysia, New Zealand and Canada	Low Indoor air pollution Low Water scarcity index Low Ozone Exceedance High Critical habitat protection High Agricultural Subsidies*	81.4%
C3 (46 countries): Majority of Central Asia and Middle East, some countries of Africa and Latin America, Brunei, Solomon Islands and Australia	Low Indoor air pollution Low Water scarcity index Low Water quality index* High CO_2 emissions*	89.1%
C4 (27 countries): The most part of Asia (Southern, South-eastern and Eastern), seven countries of Latin America, Fiji and Namibia	High Indoor air pollution* Low Disability Life Adjusted Years	48.6%
C5 (5 countries): Kuwait, Qatar, Saudi Arabia, United Arab Emirates and Libya	Low Indoor air pollution High Water Scarcity index* High Agricultural water intensity*	100%
C6 (2 countries): Bahrain and Singapore	Low Indoor air pollution Low Water Scarcity index High SO_2 emissions* High NOx emissions* High NMVOC emissions*	100%
C7 (4 countries): Angola, Brazil, Democratic Republic of Congo and United States of America	Low Indoor air pollution Low Water Scarcity index High Ozone Exceedance*	75.0%

*bad characteristics from the environmental point of view

k. The undesirable indicators that were transformed according with this procedure are signalled with an asterisk in Table 5.3. For variables with negative values in the EPI indicators (i.e., indicators O17 and O18 in Table 5.3) we made a translation of the data by summing the absolute value of the most negative observation.

Table 5.3 shows the descriptive statistics, after the transformations above mentioned, for each of the 25 indicators of the sample analyzed, consisting of 163 countries worldwide.

The next step of the analysis focused on the performance assessment of countries within each of the four largest clusters. In this section we will only present the results obtained for cluster C2 to illustrate the potential of the

Table 5.3: Variables used in the construction of CI

Indicators	Mean	SD	Min	Max
Disability Life Adjusted Years* (O1) [complement to 231.34]	172.93	52.21	21.34	218.34
Access to adequate sanitation (O2)	70.61	29.34	5.00	100.00
Access to water (O3)	84.97	16.89	40.00	100.00
Indoor air pollution* (O4) [complement to 95]	57.20	36.57	0.34	90.00
Outdoor air pollution - Urban Particulates* (O5) [complement to 145.91]	99.31	32.06	10.48	139.43
Ozone Exceedance* (O6) [complement to 86.14]	82.41	9.30	43.06	86.14
Non-methane volatile organic compound emissions* (O7) [complement to 57.39]	53.78	4.44	34.35	57.38
Sulfur dioxide emissions* (O8) [complement to 94.22]	90.92	7.24	52.28	94.22
Nitrogen oxides emissions* (O9) [complement to 76.86]	73.82	5.18	47.02	76.84
Water quality index (O10)	60.65	20.86	23.93	100.00
Water stress index (O11)* [complement to 68.55]	53.71	18.09	12.67	68.55
Water scarcity index (O12)* [complement to 4.87]	4.70	0.46	3.17	4.87
Biome protection (O13)	6.60	3.44	0.00	10.00
Critical habitat protection (O14)	24.56	28.65	7.14	100.00
Marine protection (O15)	118.13	289.71	0.33	1671.95
Growing stock change (O16)	95.01	15.87	65.02	147.75
Forest cover change (O17)	3.91	1.38	0.82	8.14
Marine trophic index (O18)	3.06	1.85	0.28	7.68
Trawling intensity (O19)* [complement to 100]	47.18	37.37	1.28	99.96
Agricultural water intensity* (O20) [complement to 398.91]	351.96	107.81	21.06	398.91
Agricultural subsidies* (O21) [complement to 54.54]	48.81	10.99	6.44	54.54
Pesticide regulation (O22)	13.10	8.55	1.00	22.00
Greenhouse gas emissions per capita* (O23) [complement to 45.89]	36.57	8.15	2.70	45.89
Industrial greenhouse gas emissions intensity* (O24) [complement to 290.86]	214.65	68.22	2.53	290.86
CO2 emissions per electricity generation* (O25) [complement to 1348.90]	862.30	270.25	197.17	1348.90

*non-isotonic output indicators

approach proposed for evaluating environmental performance. The results for the others three clusters are presented in Appendix C. The performance scores obtained using the DEA model (3.2) with the weight restrictions (5.1) are shown in Table 5.4. Note that the weight restrictions (5.1) generated 25 additional constraints, one for each indicator. Considering that the performance assessment was conducted within each cluster, the outputs of the artificial DMU referred in (5.1) corresponded to the average value of the output indicators of countries that belong to the cluster under assessment.

The best environmental performance of cluster C2, assessed taking into account the relative importance of the environmental indicators defined for the EPI, was observed in a country from Central America (Costa Rica). For the remaining countries, a score lower than 100% indicates that there is potential for improvement, supported by a comparison with Costa Rica, which is the country that presented the maximum score in this cluster. The ranking presents a satisfactory level of discrimination between the countries, which was possible due to the imposition of weight restrictions in the DEA model.

As the performance assessment conducted in our study used a system of weights that mimics the one used in the construction of the EPI, it is possible to assess the robustness of the new approach proposed in this chapter by comparing its results with the EPI ranking. Table 5.4 presents the scores and the rank position of countries obtained using DEA and the EPI methodologies. Note that while the EPI evaluated the overall set of countries (163 countries), we evaluated the performance within clusters, so the comparison between the rank positions cannot be direct. For the countries in cluster C2 (43 countries), the Spearman's rank correlation coefficient between the DEA and EPI rankings is 0.681. In order to test whether the observed correlation is significantly different from zero we performed the Spearman's rank correlation test. The p-value of the test is close to zero (p-value=0.0000), so the null hypothesis that there is no relation between the results of the approaches was rejected. The correlation analysis for the other three clusters was also tested, and for all clusters were found a significant positive correlation between the approaches. These results are presented in Appendix C.

Although the EPI and DEA methodologies use similar indicators and weighting systems, differences in results would be expected as the data used in the two approaches were subject to different normalization procedures and treatment of extreme values and outliers. In the construction of the EPI

Table 5.4: Environmental performance scores and ranks for countries of cluster C2

Country	DEA		EPI	
	Score	Rank	Score	Rank
Costa Rica	100%	1	86.4%	3
Ecuador	98.8%	2	69.3%	30
Iceland	98.7%	3	93.5%	1
France	97.9%	4	78.2%	7
Colombia	97.6%	5	76.8%	10
Germany	97.4%	6	73.2%	17
Portugal	97.3%	7	73.0%	19
Sweden	96.0%	8	86.0%	4
Italy	95.8%	9	73.1%	18
Dominican Republic	95.7%	10	68.4%	36
United Kingdom	94.7%	11	74.2%	14
Spain	94.7%	12	70.6%	25
Cuba	94.5%	13	78.1%	9
New Zealand	93.4%	14	73.4%	15
Panama	93.0%	15	71.4%	24
Norway	92.9%	16	81.1%	5
Japan	92.7%	17	72.5%	20
Switzerland	91.7%	18	89.1%	2
Latvia	91.5%	19	72.5%	21
Finland	91.2%	20	74.7%	12
Lithuania	90.6%	21	68.3%	37
Denmark	90.5%	22	69.2%	32
Romania	90.3%	23	67.0%	45
Canada	89.8%	24	66.4%	46
Croatia	89.7%	25	68.7%	35
Ireland	89.4%	26	67.1%	44
Austria	89.3%	27	78.1%	8
Malaysia	89.0	28	65.0%	54
Russia	88.9%	29	61.2%	69
Netherlands	88.0%	30	66.4%	47
Slovakia	87.2%	31	74.5%	13
Hungary	86.9%	32	69.1%	33
Bulgaria	86.9%	33	62.5%	65
Poland	86.8%	34	63.1%	63
Estonia	86.8%	35	63.8%	57
Slovenia	86.6%	36	65.0%	55
Greece	86.4%	37	60.9%	71
Czech Repuplic	85.9%	38	71.6%	22
Malta	85.5%	39	76.3%	11
South Korea	84.9%	40	57.0%	94
Belgium	83.9%	41	58.1%	88
Luxembourg	81.5%	42	67.8%	41
Cyprus	78.4%	43	56.3%	96

the raw data is transformed in a proximity-to-target value that is calculated based on the gap between the values of the indicators for each country and a target previously identified. This transformation works as a normalization process that converts all indicators to comparable measurement scales. In addition, for some indicators, a logarithmic transformation is employed to increase discrimination, and a winsorization process is used to trim the tails of distributions that presented extreme values or outliers. In our approach the performance scores were obtained based on the raw data. The DEA approach does note requires any normalization of the indicators values prior to the assessment.

In order to analyze the strengths and weaknesses of each country we relaxed the fixed weight restrictions, allowing a flexibility of $c = 0.4$ around the indicator weight w, as shown in expression (5.2). Different levels of flexibility (c) were tested before choosing $c = 0.4$. Allowing more flexibility (i.e., for higher values of c), several countries were able to obtain the maximum score, so the level of discrimination for the performance assessment would not be satisfactory. On the other hand, for lower values of c, the weights selected by the countries would become similar, such that it would be difficult to identify the categories of indicators for which the countries are doing better. Through the virtual weights chosen by each country, within the limits allowed, we are able to identify the areas in which countries are specialized and have better environmental performance. Figure 5.1 shows the results obtained for Ireland, the country selected for illustration. The pie chart (a) shows the virtual weight by category of indicators resulting from the use of the fixed weights w specified by the EPI assessment. The pie chart (b) has the virtual weights selected by Ireland for each category of indicators, given the boundaries of flexibility allowed ($c = 0.4$). Allowing this weight flexibility Ireland obtained a score of 95.5%.

Ireland selected lower virtual weights for the following categories: *Environmental burden of disease*, *Climate Change* and *Biodiversity & Habitat*. These are the categories in which Ireland needs to improve its performance

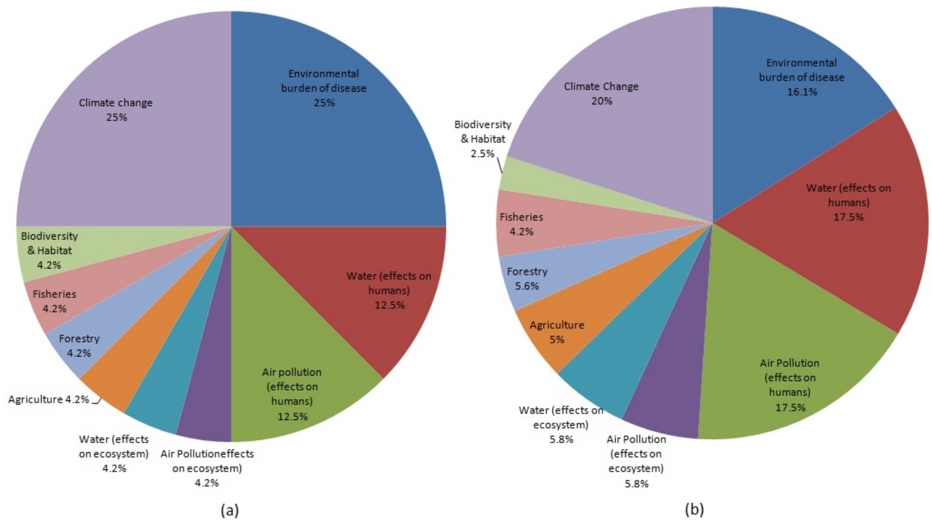


Figure 5.1: (a) Fixed weights for EPI categories and (b) contributions of the categories to the CI for Ireland

the most. On the other hand, Ireland selected higher virtual weights for the categories *Water (effects both on the humans and ecosystem)*, *Air pollution (effects both on humans and humans)*, *Agriculture* and *Forestry*. This provides evidence that Ireland is specialized in these categories. In particular for the categories related to *Water* and *Air pollution*, Ireland selected the maximum possible weight. Note that most countries from cluster C2 have good performance in indicators related to these categories, as shown by the results of the decision tree analysis reported in Table 5.2.

This methodology can also be used to point out, for a country with poor performance, the peers with a similar environmental profile that it should look to improve its performance for each indicator. Table 5.5 presents the values of the indicators of Ireland and its peers. The values of the λ_j provide an indication of the degree of similarity between Ireland (country under assessment) and each peer.

Based on the values presented by the peer countries it is possible to identify where the best practice examples for Ireland can be found for each of the

Table 5.5: Peers for Ireland

Categories	Indicators	Ireland	Peers France λ=0.689	Peers Iceland λ=0.255	Peers Costa Rica λ=0.056
Environmental burden of disease	Disability Life Adjusted Years* (O1)	215.34	214	**218.34**	211
Water (effects on humans)	Access to adequate sanitation (O2)	100	100	100	96
	Access to drinking water (O3)	100	100	100	98
Air Pollution (effects on humans)	Indoor air pollution* (O4)	90	90	90	82.26
	Outdoor air pollution - Urban Particulates* (O5)	130.32	**132.43**	127.66	109.66
Air Pollution (effects on ecosystem)	Ozone Exceedance* (O6)	86.13	85.11	86.14	86.14
	Non-methane volatile organic compound emissions* (O7)	56.67	52.94	55.05	55.67
	Sulfur dioxide emissions* (O8)	93.40	93.31	73.57	**93.65**
	Nitrogen oxides emissions* (O9)	75.07	74.23	67.90	74.59
Water (effects on ecosystem)	Water quality index (O10)	91.91	86.51	**100**	47.72
	Water stress index* (O11)	68.55	60.16	67.63	68.55
	Water scarcity index* (O12)	4.87	4.87	4.87	4.87
Biodiversity & Habitat	Biome protection (O13)	0.93	**10**	8.73	**10**
	Critical habitat protection (O14)	7.14	50	7.14	**75**
	Marine protection (O15)	0.38	**60.53**	38.53	56.27
Forestry	Growing stock change (O16)	109.36	109.36	**111.11**	100.4
	Forest cover change (O17)	6.12	4.52	**8.12**	4.32
Fisheries	Marine trophic index (O18)	4.43	4.75	3.29	**6.66**
	Trawling intensity* (O19)	39.01	75.2	46.51	**98.25**
Agriculture	Agricultural water intensity* (O20)	398.91	396.99	398.91	397.64
	Agricultural subsidies* (O21)	35.87	41.94	6.44	**54.54**
	Pesticide regulation (O22)	20	**21**	20	18
Climate Change	Greenhouse gas emissions per capita* (O23)	30.23	36.66	**45.89**	43.89
	Industrial greenhouse gas emissions intensity* (O24)	222.62	202.93	208.96	202.82
	CO_2 emissions per electricity generation* (O25)	845.15	1258.8	**1347.55**	1277.04

*non-isotonic output indicators

25 indicators. For example, considering specifically the output indicator *Disability Life Adjusted Years* (O1), which is related with premature death, Ireland can improve its performance applying the policy of Iceland (peer country which presented the best value in this indicator). If we consider the indicator related to *Greenhouse gas emissions* (O23), Iceland and Costa Rica provide good examples to learn from. A similar analysis can be done for the others indicators. Table 5.5 highlights for each indicator the best value observed in the peers. The identification of best practices for each indicator can help decision makers to define good environmental policies to improve

the performance of countries. Allowing for flexibility in the choice of weights, corresponding to a value of $c = 0.4$, the countries that are examples of best practices in cluster C2 are: Colombia, Costa Rica, Dominican Republic, Ecuador, France, Germany, Iceland, Italy, Portugal and Sweden.

5.5 Conclusions

While the climate change and warming issues are usually given more emphasis by the media in environmental reports, the world is, in fact, facing several other environmental problems that threaten the wellbeing of people on a global scale. In order to effectively tackle complex environmental issues, decision makers must first understand exactly where their countries stand in terms of environmental performance, to be able to design the steps that should be taken to improve performance.

This chapter conducted a data-driven assessment, that took into account different environmental characteristics, to deliver a robust performance assessment and suggest directions for improvement.

Firstly, by employing cluster analysis we were able to group countries into relative homogeneous groups and assess the environmental performance inside these groups. This ensured that each country was compared to other countries with similar features. Moreover, using a decision tree it was possible to identify the features that the countries from the same cluster share.

Then, in order to conduct the assessment of countries' environmental performance, we applied the approach developed in section 3.3.1 for the construction of a composite indicator in the presence of undesirable outputs. The CI model was used with weight restrictions corresponding to assurance regions type I, proposed in section 3.4.2. These restrictions are able to reflect in percentage terms the relative importance of individual indicators.

The environmental performance assessment was conducted in two stages. First, using the ARI weight restrictions, we were able to create a ranking of countries, where all countries were evaluated against a unique frontier based on a common system of weights. The use of common weights enabled a fairer comparison of countries performance, as it prevented the countries to obtain a high score only due to a careful choice of weights.

In a second moment, we are concerned with performance management and its improvement. By allowing a range of flexibility for the value of the indicators' weights, it was possible to identify the environmental strengths and weaknesses of each country. The peers with similar features to the low-performing countries were also identified. These peers provide examples of good environmental practices that the countries with worse performance should follow to improve performance.

The information provided in this chapter can support decision makers in understanding where their countries stand in terms of environmental performance, and guide the definition of environmental policies leading to performance improvements.

Chapter 6

The assessment of cities' livability integrating human wellbeing and environmental impact

6.1 Introduction

The rapid growth of urban centers and the increasing demand for products and services have led to several social, economic and environmental challenges. Overcrowding, insecurity, unemployment, natural resources depletion and pollution are just a few of the factors that deter human wellbeing and environmental quality. Governments have devoted considerable attention to these issues that directly affect livability and sustainable development of cities. In this study we address the assessment of livability in European cities. The assessment of cities' livability and the evolution of their performance compared to others play an important role in urban planning and management. Such efforts are expected to lead to better standards of human wellbeing without compromising environmental sustainability in the long-term.

National and local authorities are increasingly supporting efforts to better understand the cities' progress in terms of economic development, sustainability and livability. Examples of efforts done on this direction are the reports "State of Australian Cities" by the Australian Department of Infrastructure and Transport (2012), and the "State of the English Cities" by the English Department for Communities and Local Government (2006).

While it is becoming widely accepted that livability is an increasingly important topic in social sciences, there are components of livability which are not yet sufficiently elaborated. The English Department for Communities and Local Government (2006) report claims that much work remains to develop in order to arrive at an effective method for assessing and monitoring livability. In particular, further research is needed in order to define the appropriate weights to be assigned to the components used in livability assessments.

Although these efforts to assess cities' livability are extremely useful to understand where cities stand and to provide a starting point for discussions on issues affecting urban livability, they are not effective in providing guidelines that cities should follow to improve livability.

In this context, the purpose of this chapter is to develop a methodology for conducting a fair and data-driven assessment of cities' livability taking into account several dimensions, in order to provide viable guidelines for improvement. This was achieved by using a composite indicator that, besides of providing an overall measure of performance for each city, enables benchmarking in such a way that it becomes possible to identify the strengths and weaknesses of each city, as well as the peers with similar features to the cities with worse performance. The CI used in this assessment also has the advantage to address the issue raised by the English Department for Communities and Local Government (2006) concerning the assignment of weights to the key performance indicators, as it uses an optimization process to identify the appropriate weights to be assigned to each indicator of livability.

This study involves three stages. The first consists in defining the appropriate set of indicators to assess cities' livability. The indicators are defined based on a literature review of the studies that approached the assessment of cities' livability. The model proposed extends the concept of urban livability to include a component related to environmental sustainability, which is based on indicators capable of measuring the pressures on the environment. The conceptual model proposed includes twenty four indicators, grouped in eight dimensions of livability: *Housing quality, Accessibility and Transportation, Human health, Economic development, Education* and *Culture and Leisure*, representing the human wellbeing component, and *Solid waste* and *Air pollutants* representing the environmental impact component.

The second stage applies the Directional CI model developed in section 3.3.2 to evaluate cities' livability. This model is based on a Data Envelopment Analysis model specified with a directional distance function. In addition to providing an overall measure of cities' performance covering the two components of livability, by using different values for the directional vector it is possible to identify the cities' potential for improvement considering different perspectives concerning the components of livability. The CI model also included the novel specification of weight restrictions proposed in section 3.4.2, corresponding to assurance regions type I. The weight restrictions imposed ensure that all indicators and dimensions contribute at least with a minimum weight to the cities' livability assessment.

The major advantage of using the Directional CI in the context of the assessment of cities' livability is that it can guide improvements with different livability objectives, depending on the directional vector specified. This feature enables a more detailed characterization of the cities profile and allows understanding which component is compromising livability the most. Furthermore, the Directional CI can accommodate the undesirable outputs present in this context, namely, the indicators related to the Environmental impact component, without any transformation to their original measurement scale.

The third stage involves the assessment of performance change over time. For this purpose we used the Luenberger productivity index (Fare and Grosskopf, 2005). This analysis also enabled classifying the cities in four groups, according to their evolution of performance over time: The *star* cities (cities that presented good performance in the first time period and were also able to improve the performance over time); the *rising star* cities (cities that presented poor performance in the first time period but improved performance over time); the *falling star* cities (cities that presented good performance in the first time period but declined performance over time); and, the cities that represent a *problem* (cities that presented poor performance in the first time period and also declined performance over time).

The remainder of this chapter proceeds as follows. Section 6.2 approaches the concept of livability and provides a literature review about the dimensions used to assess livability in cities. Section 6.3 presents the methodology used to select the indicators, to construct the CI and to analyse the evolution of performance over time. Section 6.4 presents the results and discussion. Finally, section 6.5 concludes.

6.2 Assessing livability

6.2.1 The concept

As pointed out by the report of the English Department for Communities and Local Government (2006), defining livability is a minefield. It is a relatively new policy area, and therefore there are competing ideas about what should be covered by this large umbrella. An easy and objective definition of livability is provided by the Merriam-Webster dictionary, where livability is defined as "suitability for human living" (Merriam-Webster, 2013). Different and more explanatory definitions can be found in the literature, e.g., Newman et al. (1996) and Newman (1999) state that livability is the human requirement for social amenity, health and wellbeing and it includes both individual and community wellbeing. In addition, livability may be related to how easy a place is to use and how safe it feels. It is an environment that is both inviting and enjoyable, where people want to live and work now and in the future (English Department for Communities and Local Government, 2006). A more complete and understandable explanation about livability is presented in the report State of Australian Cities, (Australian Department of Infrastructure and Transport, 2012), that states that livability describes the degree to which a place supports quality of life, health and wellbeing. A liveable city should be healthy, safe, harmonious, attractive and affordable. It should also have high amenity, provide good accessibility and be environmentally sustainable.

Although we can find in the literature some attempts to explain the differences between livability and quality of life, in some studies the livability concept is used interchangeably with the concept of quality of life, as some indicators are considered both in livability and quality of life assessments, e.g. Pichardo-Muiz (2011) and Morais and Camanho (2011). While the terms embody similar concepts, the distinction lies in the difference between the presence and quality of the amenities (livability) and the user experience of those amenities (quality of life). For example, for the Equity dimension, the variable "equitable distribution of amenities" would be related to livability, and the "sense of social justice, exposure to diverse ideas" would be related to quality of life (Oregon Department of Transportation, 2011).

6.2.2 The dimensions

The literature review conducted in this chapter has focused on the identification of the different dimensions of livability that are appropriate for the assessment of urban areas. This provided the foundations to construct a conceptual model composed by indicators that are able to translate as well as possible the livability concept. This conceptual model aims to provide a basis for the assessment and monitoring of cities' livability.

Newman et al. (1996), in the Australian State of Environment report, defined a conceptual model to assess sustainability. This model considers that a city can only be considered sustainable if it reduces the flow of materials crossing the city, and also ensures a minimum standard of livability. The dimensions proposed to assess livability include *Wealth and Income inequality, Unemployment, Education and training, Housing, Accessibility and urban design* and *Health*. Some years later, the Australian State of Environment report was updated by Newton et al. (2001). These authors believe that urban populations are directly affected by the quality of their immediate physical environment, so they included in the framework proposed by Newman et al. (1996) indicators related to the *Environmental health* (i.e. noise, indoor air pollutants, green spaces).

Some of the dimensions proposed in the Australian reports are also mentioned in the Livability and Quality of Life Indicators report, provided by the Oregon Department of Transportation (Oregon Department of Transportation, 2011). The proposed dimensions to assess livability described in this report are related to *Economic Development, Housing, Environmental Quality, Community Development* and *Equity.*

Another governmental effort to account for livability in urban centres was done in the State of the English Cities report (English Department for Communities and Local Government, 2006). The analytical framework proposed in this report covers four themes, namely, *Environmental Quality, Physical Place Quality, Functional Place Quality* and *Safer Places.* Although the last component, related to crime and safety, is not explicitly mentioned in the frameworks proposed by Newman et al. (1996) and Newton et al. (2001), it composes an important part of the wider livability agenda in the State of the English Cities report.

In 2010, the Australian Government, through the Department of Infrastructure and Transport, developed the State of Australian Cities report, that has been yearly updated since then (Australian Department of Infrastructure and Transport, 2010). The report aims to highlight emerging trends and issues to promote discussion on the management of growth and change in the Australian major urban centres. In the last version, published in 2012, the assessment of cities' livability included, besides of the dimensions proposed in Newman et al. (1996) and Newton et al. (2001), components related to *Equality, Safety, Affordability* and *Community wellbeing.*

6.3 Methodology

6.3.1 Selection of indicators

The range of dimensions used in this chapter to assess the livability of European cities were defined based on a literature review. We followed mainly the approaches proposed in the Australian State of the Environment report, by Newman et al. (1996) and later updated by Newton et al. (2001). We also considered the State of the English Cities (English Department for Communities and Local Government, 2006) and State of Australian Cities reports (Australian Department of Infrastructure and Transport, 2012), that include some dimensions that were not included in Newman et al. (1996) and Newton et al. (2001), as discussed in the Section 6.2.2. As defined in the report of the English Department for Communities and Local Government (2006), livable places are places where people want to live and work now and in the future. This definition recognizes an implicit link between livability and sustainability. Thus, places with good livability must not only be economically and socially successful, but also need to have a low environmental impact. Low levels of environmental impact is a key factor to achieve sustainability. Thus, the performance assessment of cities reported in this chapter also included indicators able to assess the environmental impact of human living. These indicators are included in the dimensions related to *Solid waste* and *Air pollutants.* They represent a broad concept of livability, that addresses also the cities' sustainability. Our conceptual model accounts for the current ability of a city to offer a place suitable for human living without compromising an identical context for the next generations.

The conceptual model proposed is defined by eight dimensions of livability: *Housing quality, Accessibility and Transportation, Human health, Economic development, Education* and *Culture and Leisure*, representing the human wellbeing component, and *Solid waste* and *Air pollutants* representing the environmental impact component. These eight dimensions are further subdivided into twenty four sub-themes which represent our range of indicators, as shown in Table 6.1. These indicators are used as a basis for the assessment of livability in European cities presented in this chapter. Therefore, first we defined the dimensions of livability that should compose the conceptual model, and, in a second moment, we selected a range of available indicators that could represent each livability dimension.

Initially, the conceptual model also included the indicators *Percentage of the urban waste water load (in population equivalents) treated according to the applicable standard*, *Proportion of residents exposed to traffic noise at day time and at night time*, and *CO_2 emissions per capita*, but due to lack of data available, these indicators could not be considered.

Table 6.1: Selected indicators to assess cities' livability

Livability components	Dimensions	Indicators
Human wellbeing	Housing quality	Average living area per person (m^2) Proportion of households living in owned dwellings in cities
	Accessibility and Transportation	Multimodal accessibility (EU27=100) Share of journeys to work not done by car Length of public transport network per inhabitant
	Human health	Life expectancy Infant Survival rate (per 1000 live births) Available hospital beds in cities (per 1000 inhabitants)
	Economic development	Employment per 100 of residents aged 15-64 GDP per head Median disposable annual household income % of the households receiving less than half of the national average household income Population per recorded crime
	Education	Proportion of students completing their compulsory education Students in upper and further education per 1000 resident pop.
	Culture and Leisure	Annual cinema attendance per resident Annual visitors to museums per resident Number of libraries per 1000 residents Green space to which the public has access, per capita
Environmental impact	Solid waste	Collected solid waste - tonnes per inhabitant and year Proportion of solid processed by landfill
	Air Pollutants	Accumulated ozone concentration in excess 70 microgram/m^3 Annual average concentration of NO_2 Annual average concentration of PM10

The data used in this study was provided by the European Union through the Urban Audit program. This program provides indicators related to eight dimensions corresponding to demography, social and economic aspects, civic involvement, education, environment, transport and travel, culture and leisure, and innovation and technology for European cities. The database of the Urban Audit project contains public data for a large number of European cities, for different periods of time. Data from two periods were used in this study, corresponding to the years 2003 to 2006 and 2007 to 2009.

6.3.2 Composite indicator model

The composite indicator used in this chapter for the assessment of livability in European cities is derived from a DEA model specified using a directional distance function. It follows the direct approach to deal with the undesirable outputs, discussed in section 3.3.2. The direct approach allows treating the outputs in their original form, that is, without requiring any modification to the measurement scale of the undesirable factors. The model used in this chapter corresponds to formulation (3.6) of section 3.3.2.

6.3.2.1 Directional vectors and objectives regarding Human wellbeing and Environmental impact

In this study, we used three different directional vectors in the Directional CI model (3.6), depending on the objective pursued. First, we are interested in assessing the livability of cities taking into account both components of livability: human wellbeing and environmental impact. So, we specified a vector that is able to account for improvements in both components: $g = (g_y, -g_b) = (y_{rj_0}, -b_{kj_0})$. By setting the directional vector as $g = (y_{rj_0}, -b_{kj_0})$ in model (3.6), i.e. the current value of the outputs for the DMU under assessment, it is possible to simultaneously expand the desirable outputs and contract the undesirable outputs through a path that allows proportional interpretation of improvements. This means that the direction of the projection to the frontier depends on the individual outputs values of each DMU. Note that the directional distance models are units invariant when the directional vector is specified as being the observed value of the desirable and undesirable outputs of the DMU under assessment.

In a second moment, we are interested in assessing the cities' potential for improvement in each component of livability. This assessment allows to understand, for those cities that did not achieve the maximum performance in the livability assessment, which component is compromising the livability the most. In order to assess the extent to which the human wellbeing indicators could be proportionally increased whilst keeping the environmental impact indicators unchanged, we specified the directional vector as equal to $g = (g_y, -g_b) = (y_{rj_0}, 0)$ in model (3.6). In this case, the factor β, corresponds to the potential for improvement only in the indicators related to human wellbeing. Similarly, in order to assess the cities' potential for improvement in the indicators related to the environmental impact component, the directional vector was specified as $g = (g_y, -g_b) = (0, -b_{kj_0})$ in model (3.6). This vector allows to assess the proportion by which all environmental indicators could be improved, while maintaining the human wellbeing indicators fixed.

Figure 6.1 shows the production frontier that would be obtained, using the Directional CI model (3.6), for a small example involving 3 DMUS (A, B and C), a desirable and an undesirable output. The production possibility set is bounded by the segments linking OABB'O. The three different vectors used in this study are also illustrated in Figure 6.1.

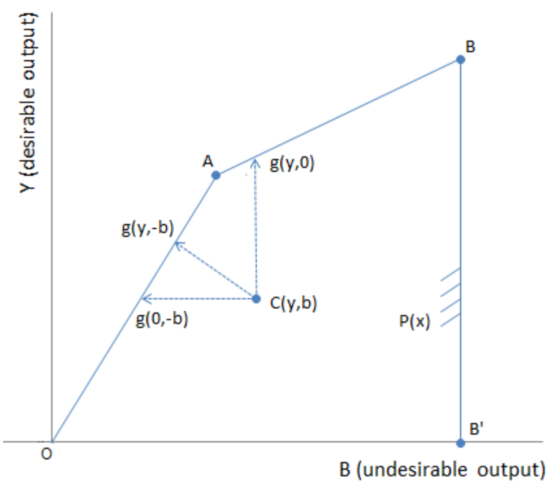


Figure 6.1: Production possibility set and the different directional vectors used in this study

Note that, while the directional vector $g = (y_{rj_0}, -b_{kj_0})$ allows to, simultaneously, identify the potential for improvements in both outputs, the other two vectors, $g = (y_{rj_0}, 0)$ and $g = (0, -b_{kj_0})$, focus exclusively on improvements to desirable and undesirable outputs, respectively, i.e., they allow to identify the potential for improvement focusing separately on each component of livability.

6.3.2.2 Incorporating information of the relative importance of the individual indicators

In the context of this chapter we decided to reduce the flexibility in the choice of weights inherent of DEA technique. The weight restrictions are imposed to the dual of model (3.6), shown in formulation (3.8).

Two types of weight restrictions were imposed. First, we imposed that each indicator should weight at least a minimum (1%) in the livability assessment, as shown in formulation (6.1). These restrictions are imposed in order to avoid having indicators that do not contribute to the composite indicator.

$$\begin{cases} \dfrac{u_r\ \bar{y}_r}{\sum_{r'=1}^{s} u_{r'}\ \bar{y}_{r'} + \sum_{k'=1}^{l} p_{k'}\ \bar{b}_{k'}} \geq 0.01 & r = 1, ..., s \\ \dfrac{p_k\ \bar{b}_k}{\sum_{r'=1}^{s} u_{r'}\ \bar{y}_{r'} + \sum_{k'=1}^{l} p_{k'}\ \bar{b}_{k'}} \geq 0.01 & k = 1, ..., l \end{cases} \tag{6.1}$$

A second type of weight restrictions was imposed to each dimension. We assume that good livability implies good standards in all dimensions simultaneously. Therefore, it was imposed that each dimension must have its weight varying between 10% and 15%. These restrictions aim to ensure that the weights are homogeneously distributed among the 8 dimensions. The restrictions to the dimension $D_z, z = 1, ..., 8$, can be seen in (6.2).

$$0.10 \leq \frac{\sum_{r \in D_z} u_r \ \bar{y}_r + \sum_{k \in D_z} p_k \ \bar{b}_k}{\sum_{r=1}^{s} u_r \ \bar{y}_r + \sum_{k=1}^{l} p_k \ \bar{b}_k} \leq 0.15 \tag{6.2}$$

Different levels of flexibility were tested before setting the values for these restrictions. By allowing more flexibility, several cities were able to obtain the maximum composite indicator score, and thus the level of discrimination for the performance assessment would not be satisfactory. On the other hand, by reducing the flexibility, the weights assigned to the indicators would become identical for all cities, such that it would be difficult to identify the dimensions and indicators that represent the strengths or weaknesses of each city.

These weight restrictions correspond to the novel specification of ARI, proposed in section 3.4.2. Recall that the advantage of using this specification of weight restrictions is that they allow setting the weight bounds in percentage terms. The use of the "artificial" DMU makes the restrictions work as ARIs, avoiding the problems associated with the specification of DMU-specific virtual weight restrictions.

The assessment of livability in European cities was conducted using model (3.8) with the weight restrictions shown in (6.1) and (6.2). These weight restrictions ensure that the cities are being evaluated under similar conditions. The use of a similar system of weights enables a fairer comparison of cities livability, as the discrimination of the performance assessment is improved by preventing the cities from obtaining a high score only owing to a judicious choice of weights (Despotis, 2002; Angulo-Meza and Lins, 2002). The dual of model (3.8) complemented with the restrictions shown in (6.1) and (6.2), is shown in the Appendix D.

6.3.3 Change of performance over time

In order to assess the change in performance over time, we used the Luenberger productivity index, introduced in section 4.3.3. According to Fare and Grosskopf (2005), the Luenberger productivity index is appropriate to analyse productivity change in assessments conducted using a directional distance function.

The Luenberger productivity index, proposed by Chambers (1996), for an assessment involving inputs x_{ij} $(i = 1, ..., m)$, desirable outputs y_{rj} $(r = 1, ..., s)$ and undesirable outputs b_{kj} $(k = 1, ..., l)$, is defined as follows:

$$\begin{aligned} L^{t,t+1} = \tfrac{1}{2}\big[&\vec{D}^{t+1}(x^t,y^t,b^t;y^t,-b^t)-\vec{D}^{t+1}(x^{t+1},y^{t+1},b^{t+1};y^{t+1},-b^{t+1}) \\ &+\vec{D}^t(x^t,y^t,b^t;y^t,-b^t)-\vec{D}^t(x^{t+1},y^{t+1},b^{t+1};y^{t+1},-b^{t+1})\big] \end{aligned} \tag{6.3}$$

This index can be addictively decomposed in two components: efficiency change and technological change, as shown in (6.4) and (6.5), respectively.

$$LEC^{t,t+1} = \vec{D}^t(x^t,y^t,b^t;y,-b)-\vec{D}^{t+1}(x^{t+1},y^{t+1},b^{t+1};y^{t+1},-b^{t+1}) \tag{6.4}$$

$$\begin{aligned} LTC^{t,t+1} = \tfrac{1}{2}\big[&\vec{D}^{t+1}(x^{t+1},y^{t+1},b^{t+1};y^{t+1},-b^{t+1})-\vec{D}^t(x^{t+1},y^{t+1},b^{t+1};y^{t+1},-b^{t+1}) \\ &+\vec{D}^{t+1}(x^t,y^t,b^t;y^t,-b^t)-\vec{D}^t(x^t,y^t,b^t;y^t,-b^t)\big] \end{aligned} \tag{6.5}$$

Recall that the Luenberger productivity index and its components signal improvements in productivity when values are greater than zero, and decline in productivity when values are less than zero. The change in relative efficiency between the periods t and $t+1$ can be interpreted as the change in the distance between the observed production and the maximum potential production. Improvements in the efficiency change component are evidence of catching up to the frontier. The technological change component captures the shift in technology between the two time periods assessed. Improvements in this component are evidence of innovation.

Following Fare et al. (1994b), we are able to determine which DMUs are the "innovators", i.e., which of them are responsible for the shifts in the frontier towards more productive levels. In addition to presenting positive levels of technological change ($LTC_t^{t+1} > 0$), the DMUs should be located on the frontier in the second time period ($t+1$), and this production point must be above the frontier of the first time period (t). In the context of the assessment of productivity change using the Luenberger index, these three characteristics can be stated as follows:

$$\begin{aligned} LTC_t^{t+1} &> 0 \\ \vec{D}_o^{t+1}(x^{t+1}, y^{t+1}, b^{t+1}; y^{t+1}, -b^{t+1}) &= 0 \\ \vec{D}_o^{t}(x^{t+1}, y^{t+1}, b^{t+1}; y^{t+1}, -b^{t+1}) &< 0 \end{aligned} \tag{6.6}$$

6.4 Results and discussion

6.4.1 Data

In order to improve the robustness of the results, the study only included indicators with a minimum of 60% valid observations throughout all cities in the sample in both time periods analysed. Furthermore, the final sample only retained the cities with data for at least one indicator in each dimension. As a result, the sample used to conduct this study is composed by 120 cities.

After excluding the indicators and cities with less than 60% of valid data and the cities without information of at least one indicator in all dimensions, 23% of the data were still missing. We used a proxy variable that corresponds to the values of these indicators in the previous time period to replace the missing values in each city. This procedure replaced 12.6% of the missing values.

Finally, in order to conduct the DEA assessment, the remaining missing values, 10.4%, were substituted by the minimum (maximum) value observed in each desirable (undesirable) indicator. This approach for handling the issue of missing values in DEA applications ensures that the performance of the DMUs is not improved by the lack of data available.

6.4.2 Results from the assessment of cities' livability

The performance assessment conducted in this section allows to identify the best performing cities of the analysed sample. It can also be used to identify the areas in which intervention is most needed to increase cities' livability. The data used in this section are related to the time period from 2007-2009.

The results presented in this section were obtained using the DEA-based CI model presented in (3.8), with the 24 output indicators described in section 6.3.2, subject to the weight restrictions presented in (6.1) and (6.2).

The results showed that the performance scores for the sample analyzed vary from 0 (best) to 0.218 (worst), and the average score is 0.066. Among the 120 European cities included in the analysis, the ones that reached the highest score (34 cities) are listed in Table 6.2. We can see that the majority of these cities belongs to Germany (12). It is worth mentioning that this country had also the largest number of cities included in the analysis (39). The whole set of cities analysed and their respective performance scores can be seen in Appendix E (column named "Livability 2007-2009").

Using the virtual weights chosen by each city within the limits allowed, we were able to identify the areas in which cities have better performance. Figure 6.2 shows the weights used in the assessment of Berlin, the city selected for illustration. Berlin obtained a composite indicator score equal to 0.086, and assigned higher virtual weights to the dimensions related to *Housing quality*, *Accessibility and Transportation*, *Human health* and *Eco-*

Table 6.2: Cities that reached the best performance score

City (country)	City (country)	City (country)
Brugge (Belgium)	Koblenz (Germany)	Lisboa (Portugal)
Burgas (Bulgaria)	Mainz (Germany)	Bern (Switzerland)
Ruse (Bulgaria)	Nurnberg (Germany)	Zurich (Switzerland)
Sofia (Bulgaria)	Potsdam (Germany)	Banska Bystrica (Slovakia)
Tallinn (Estonia)	Regensburg (Germany)	PreSov (Slovakia)
Oulu (Finland)	Schwerin (Germany)	Trencin (Slovakia)
Tampere (Finland)	Weimar (Germany)	Pamplona/Iruna (Spain)
Bochum (Germany)	Bergen (Norge)	Santiago de Compostela (Spain)
Bonn (Germany)	Stavanger (Norge)	Goteborg (Sweden)
Darmstadt (Germany)	Tromso (Norge)	Stockholm (Sweden)
Dusseldorf (Germany)	Konin (Poland)	
Frankfurt am Main (Germany)	Funchal (Portugal)	

nomic development. This provides evidence that Berlin is well developed in these areas. On the other hand, Berlin selected lower virtual weights for the following dimensions: *Education*, *Culture and Leisure*, *Solid Waste* and *Air Pollution*. These are the dimensions in which Berlin could improve its performance the most. It is also possible to see in Figure 6.2 the weights allocated to some indicators inside each dimension. For example, in the dimension related to *Economic development*, Berlin allocated the largest possible weight to the indicator *Percentage of the households receiving more than half of the national average income*, which reveals a particularly good performance in this feature.

This methodology can also be used to point out, for a city with low performance, the peers with a similar livability profile that the city should examine in order to learn with the examples of best practices. The peers of Berlin are Stockholm, Mainz, Weimar, Lisbon and Tampere. It is also interesting to know in which dimensions the peer cities of Berlin are allocating the largest amount of weight, i.e. the features in which their performance is particularly good. Figure 6.3 presents the weights assigned to each dimension by Berlin and its peers. The value of λ_j at the optimal solution to the Directional CI model (3.6) is also presented between brackets in Figure 6.3. It provides an indication of the degree of similarity between Berlin (city under assessment) and each peer. For example, considering specifically the

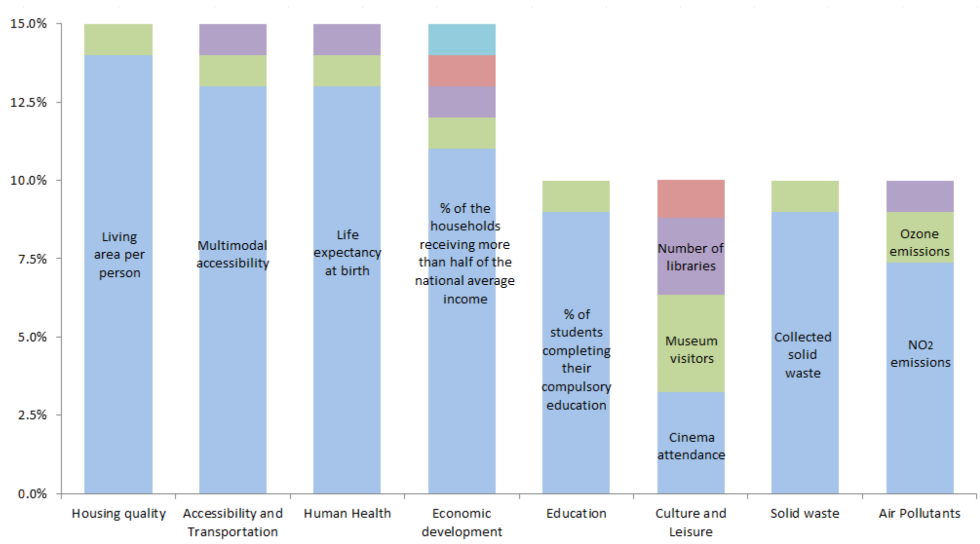


Figure 6.2: Virtual weights selected by Berlin in each dimension

dimension *Air Pollutants*, Berlin can improve its performance by applying the policy of Stockholm, Weimar and Tampere (peer cities that presented good performance in this dimension, as they assigned to it the maximum possible weight: 0.15). If we consider the dimension related to *Education*, Stockholm, Lisbon and Tampere provide good examples to learn from. A similar analysis can be carried out for the other dimensions.

6.4.3 Cities' potential for improvement in each component of livability

As explained in section 6.3.2.1, by assuming different scenarios for the directional vector g in the CI model (3.6), it is possible to calculate the potential for improvement for each component of livability, representing different priorities concerning human wellbeing and environmental impact of European cities. While the overall indicator of livability points out how much a city could improve all indicators simultaneously, the assessment that focuses specifically on each component indicates how much a city can improve the indicators related to that component. This assessment also enables a more

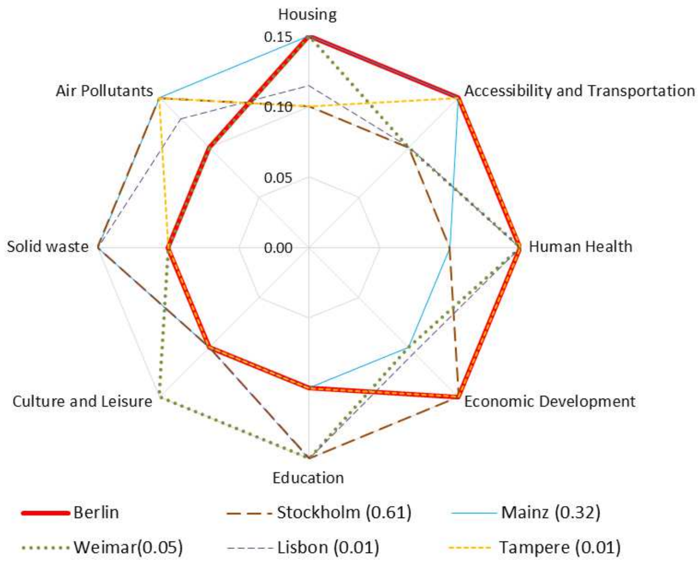


Figure 6.3: Weights assigned by Berlin and its peers to livability dimensions

detailed characterization of cities profile and allows to understand which component (human wellbeing or environmental impact) is compromising livability the most.

As pointed out by Kuosmanen and Matin (2011), the choice of the directional vector can influence the distance of the DMUs to the frontier and the benchmarks, but it does not influence the shape of the frontier. This means that the 34 cities that reached the best livability score will have the same performance ($\beta = 0$) in the assessments corresponding to different livability objectives reflected by the specification of different directional vectors. Therefore, the discussion included in this section involves only the 86 cities that did not reach the best livability score.

The performance score obtained, in each component of livability, for these 86 cities varies from 0 to 0.272 in the human wellbeing component and from 0 to 0.913 in the environmental impact component. Note that the environmental component has a larger range of values because the proportional improvement is only sought in 5 indicators, whereas human wellbeing requires improvement in 19 indicators.

It is possible to see a strong correlation between the scores of the two components ($r = 0.894$). A significant relationship can also be seen between the livability scores and the performance obtained for each component of livability considered individually.

The average value of the scores obtained in the livability assessment for all cities is equal to 0.066. In the assessments focusing on the environmental impact and human wellbeing components separately, the average is equal to 0.325 and 0.081, respectively. Thus, the average value of the environmental impact component is 4.944 times superior than the average value of the livability scores. This means that, by focusing the improvements only in the environmental impact component, there is a scope for improvement 4.944 times superior than if we consider the improvements in all indicators simultaneously. On the other hand, by focusing on the improvements exclusively in the human wellbeing component, the scope for improvement is, on average, 1.231 superior to what would be feasible considering improvements in all indicators simultaneously. This analysis allowed to identify the cities that have an unbalanced potential for improvement among the two components. A city that has an environmental impact component superior to 4.944 times the value of its own livability score is considered as having a scope for improvement in this component larger than what it would be expected. Similarly, a city is considered to have a larger scope for improvement than the expected in the human wellbeing component, when its score is larger than 1.231 times its own livability score.

Lets consider, for example, the city of Dortmund, in Germany, that obtained a score equal to 0.140 in the livability assessment. Dortmund has scores equal to 0.698 and 0.166 in the evaluation focusing on the environmental impact and human wellbeing components, respectively. This means that, Dortmund could improve the indicators related to environmental impact by the ratio of 0.698 while keeping unchanged the indicators related to human wellbeing. Similarly, keeping the indicators related to environmental impact unchanged, Dortmund has the potential to improve the indicators related to human wellbeing by the ratio of 0.166.

Comparing these scores with the expected values ($0.140 \times 4.944 = 0.692$ for the environmental impact component, and $0.140 \times 1.231 = 0.172$ for the human wellbeing component) we can conclude that Dortmund has a potential for improvement in the environmental impact component beyond expectations ($0.698 > 0.692$) and the reverse occurs for the human wellbeing component ($0.166 < 0.172$, meaning a potential for improvements below expectations). This means that Dortmund has an unbalanced potential for improvement among the livability components, and the largest potential for improvement is associated with the environmental component.

The scores regarding the human wellbeing and environmental impact components for all cities for the time period 2007-2009 are reported in Appendix E. Among the 120 cities analyzed, 78 cities presented an unbalanced potential for improvement among the two components. This is signed in Appendix E using the symbol "▲" next to the score of the component with a potential for improvement larger than average. Overall, 48 cities have larger scope for improvement in the environmental impact component, and 28 in the human wellbeing component. In addition, it is worth noting that two cities (Olsztyn and Rzeszow, from Poland) presented a potential for improvement above their expected value in both components.

6.4.4 Assessing performance change over time

This section presents the results of the assessment regarding the change in cities' livability over time, considering the two time periods: 2003-2006 and

2007-2009. The assessment was conducted using the Luenberger productivity index, explained in Section 6.3.3. In order to estimate the directional distance functions of the Luenberger productivity index we used the model (3.8) specified with a directional vector equal to $g = (y_{rj_0}, -b_{kj_0})$, subject to the weight restrictions presented in (6.1) and (6.2).

Figure 6.4 presents, in the X-axis, the livability scores obtained by the cities in the first time period (2003-2006) and, in the Y-axis, the Luenberger productivity index, which reflects the changes in performance between the period 2003-2006 and the period 2007-2009. Points located above the horizontal axis represent cities that increased performance over time, whereas cities located below declined performance over time. In addition, points located on the left of the vertical segment, corresponding to the median livability score ($\beta = 0.070$), represent cities that presented good performance in the first period analysed, whereas to the right of this segment are cities that presented poor performance. This analysis enabled classifying the cities in four groups, according to their original values of livability and evolution over time.

The cities that belong to the first group, corresponding to the North-West quadrant, are considered *stars* (32 cities). They presented good performance in the first time period and were still able to improve the performance over time (Luenberger productivity index larger than zero). The cities in the second group (North-East quadrant) are considered *rising stars* (48 cities). They presented bad performance in the first time period but improved performance over time. The third group (South-West quadrant), named *falling star* (28 cities), is composed by cities with good performance in the first time period but that declined performance over time. The last group, named *problem* (South-West quadrant), corresponds to cities that presented bad performance in the first time period and also declined performance over time (12 cities).

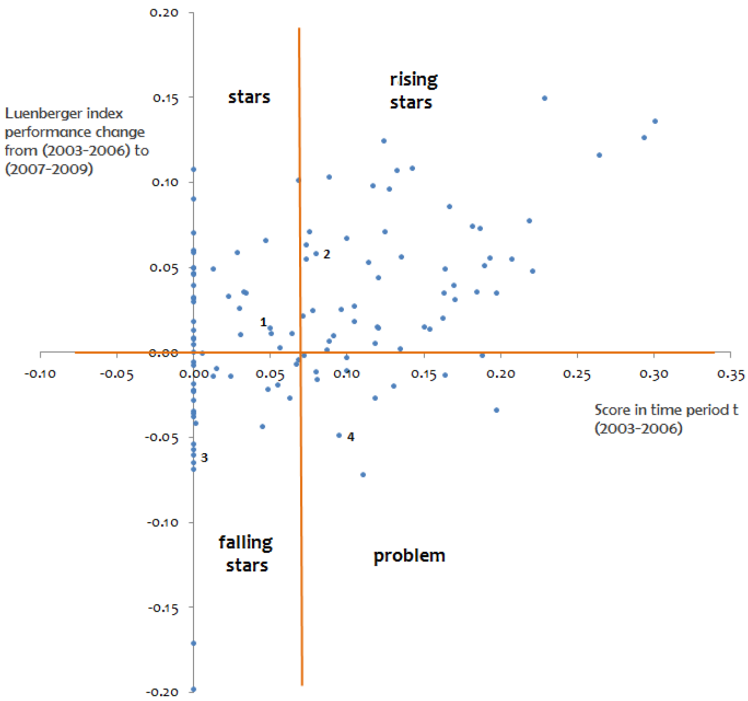


Figure 6.4: Cities' performance over time

Munich is an example of a *star* city. In the first time period, it had a level of livability among the 50% best scores ($\beta = 0.017$) and a positive Luenberger productivity index ($L^{t,t+1} = 0.014$). In Table 6.3, we can see that the Munich' productivity improvement is due to an improvement in the efficiency change component, whose value is 0.033, meaning that it is closer to the best practice frontier in the second time period. In addition, we can see that Munich was projected to a part of the frontier where technological regression occurred (technological change component equal to -0.019). This means that, in the second time period, the best practices in livability of cities with a profile identical to Munich had worse livability standards than in the previous time period.

Antwerpen is an example of a city classified as *falling star*. It presented a good livability score in the first period analysed, but the Luenberger

Table 6.3: Results for the assessment of performance change over time for four cities used as examples

City	Classification	Luenberger Index	Efficiency Change	Technological Change	Score in 2003-06	Score in 2007-09
Munich (1)	star	0.014	0.033	-0.019	0.050	0.017
Madrid (2)	rising star	0.058	0.054	0.004	0.080	0.026
Antwerpen (3)	falling star	-0.060	-0.017	-0.043	0	0.017
Malmo (4)	problem	-0.048	-0.015	-0.033	0.095	0.110

productivity index was equal to -0.060. The decline in performance over time was mainly due to a decline in the technological change component ($LTC^{t,t+1} = -0.043$), meaning that the best practice frontier declined. The efficiency change component also declined ($LEC^{t,t+1} = -0.017$), meaning that the city became further apart from the best practice frontier. Table 6.3 also presents the detailed results obtained for Madrid and Malmo, that are examples of *rising star* and *problem* cities, respectively. The position of these cities is identified in Figure 6.4 using the labels 1, 2, 3 and 4, referring to Munich, Madrid, Antwerpen and Malmo, respectively.

It is worth noting that some cities that were located in the falling star group have good livability scores in the assessments conducted in both time periods. For these cities, the negative value of the Luenberger productivity index was due to the regression in the frontier. The financial crisis that affected the European countries since the beginning of 2008 may be a cause of this frontier decline. Stockholm is an example of a city in this situation: it presented the best livability score in both time periods ($\beta = 0$), but has a Luenberger index equal to -0.036.

The value of the Luenberger index and its components (efficiency change and technological change), as well as the classification of all cities in the four groups (*star, rising star, falling star* and *problem*) is presented in Appendix E.

The results of the Luenberger productivity index range from -0.198 to 0.150

(0.021 on average). Las Palmas, in Spain, presented the highest improvement in productivity over time (0.150), this progress is mainly due to improvements in efficiency. Las Palmas has one of the highest values of the efficiency change component in the sample analysed. The results of the decomposition of the Luenberger productivity index for all cities shows that, although the average value of the technological change component is positive (0.009), the efficiency change component is the main source of performance improvement, with an average value of 0.012.

Table 6.4 shows that the technological progress was, on average, larger in cities classified as *stars* (0.033). Conversely, the largest technological regress, on average, is presented by the *falling star* cities (-0.014). The *falling star* group also presented the largest decline in efficiency over time (-0.026 on average). The largest efficiency improvement (0.046 on average) was presented by cities classified as *rising stars*. This group also presented the largest change in productivity, with an average value of the Luenberger index equal to 0.056.

Table 6.4: Average results for the efficiency and technological change and Luenberger index per group of cities

Cities' classification	Efficiency change Average	Technological change Average	Luenberger index Average
Stars	0.006	0.033	0.038
Rising stars	0.046	0.010	0.056
Falling stars	-0.026	-0.014	-0.040
Problem	-0.013	-0.008	-0.021

As it can be seen in Table 6.5, among the 120 cities analysed, 61 presented an improvement in efficiency over time, and the majority of them belong to the *rising star* group. In addition, 72 cities are close to regions of the frontier where progress occurred, and most of them belong to the *star* or *rising star* groups. It is also possible to see that the majority of cities (80) presented positive values of Luenberger index, suggesting improvements in productivity over time. The largest number of cities that improved productivity over time belong to the *rising star* group.

Table 6.5: Number of cities that declined, maintained or improved the levels of productivity over time, per group

Cities' classification	Number of cities	Efficiency change			Technological change			Luenberger index		
		↘	Maintain	↗	↘	Maintain	↗	↘	Maintain	↗
Stars	32	4	18	10	3	0	29	0	0	32
Rising stars	48	4	1	43	15	1	32	0	0	48
Falling stars	28	18	5	5	20	1	7	28	0	0
Problem	12	9	0	3	8	0	4	12	0	0
Total	**120**	**35**	**24**	**61**	**46**	**2**	**72**	**40**	**0**	**80**

Among the 72 cities that have a positive technological change component, we can identify the "innovators", i.e., those cities that are responsible for the shifts in the frontier towards more productive levels. Following the requirements to be considered innovative, explained in section 6.3.3, 24 cities were identified as innovators, they are presented in Table 6.6. Among these cities, 23 belong to the *star* group and 1 (Konin) to the *rising star* group.

Table 6.6: Cities considered innovators

City (country)	City (country)	City (country)
Brugge (Belgium)	Frankfurt am Main (Germany)	Lisboa (Portugal)
Burgas (Bulgaria)	Mainz (Germany)	Banska Bystrica (Slovakia)
Ruse (Bulgaria)	Potsdam (Germany)	PreSov (Slovakia)
Sofia (Bulgaria)	Weimar (Germany)	Trencin (Slovakia)
Tallinn (Estonia)	Bergen (Norge)	Pamplona/Iruna (Spain)
Tampere (Finland)	Tromso (Norge)	Santiago de Compostela (Spain)
Bochum (Germany)	Konin (Poland)	Goteborg (Sweden)
Darmstadt (Germany)	Funchal (Portugal)	Bern (Switzerland)

6.5 Conclusions

This study contributed to the development of a tool to assess the livability of cities and promote its improvement. The indicators used to measure livability were defined based on a literature review of the studies that approached cities' livability assessment. The conceptual model proposed is based on a broad concept of livability that includes social and economic aspects, as well as principles of environmental sustainability. The model involves 24 indicators grouped in 8 dimensions: *Housing quality, Accessibil-*

ity and Transportation, Human health, Economic development, Education and *Culture and Leisure*, representing the human wellbeing component, and *Solid waste* and *Air pollutants* representing the environmental impact component.

Furthermore, it was used a Data Envelopment Analysis model, specified with directional distance function, for the construction of a composite indicator to assess cities' livability. This lead to the estimation of an overall measure of livability for each city. This measure can be interpreted as the cities' potential for improvement in all livability indicators simultaneously. The specification of different directional vectors in the Directional CI model also allowed to identity the potential for improvement in specific components of livability (i.e. human wellbeing and environmental impact). The possibility to specify different directional vectors reflecting different priorities to improvements in specific dimensions represent the main methodological contribution of this chapter. Another important feature of the composite indicator used in this chapter consists in the incorporation of weight restriction that allowed to incorporate the relative importance of indicators, expressed as a percentage, using assurance regions type I.

In addition to providing an enhanced picture of livability of European cities, the model described in this chapter can be used for benchmarking purposes, as it suggests to each city with poor livability a set of peers whose practices are examples to be followed. Furthermore, the specification of the weights recurring to optimization allowed to identify the strengths and weaknesses of each city. This information can support decision makers in the definition of policies to improve the performance of cities taking into account specific features and preferences.

In the last stage, this chapter also approached the assessment of performance change over time. For this purpose we used the Luenberger productivity index, which is appropriate for assessments conducted using Data Envelopment Analysis models specified with a directional distance function. This analysis also enabled classifying the cities in four groups *star*, *rising star*, *falling star* and *problem* cities, according to their changes in performance over time. This information can support decision makers to understand the evolution of cities' livability. The decomposition of the Luenberger productivity index allowed analysing the changes in cities' performance over time as well as to identify progression or regression in the production frontier. In addition, we pointed out the cities that can be considered innovators, i.e., those cities that are responsible for the shifts in the frontier towards more productive levels.

We believe that this study enhanced the understanding of livability of European cities. Besides providing decision makers information that enables them to understand exactly where each city stands in terms of livability, the methodology used here also suggested the steps that can be taken to improve cities' performance.

Chapter 7

Conclusions

7.1 Summary and conclusions

This thesis addressed two main objectives. The first was related to the development of innovative models for the assessment and monitoring of performance in the presence of undesirable outputs using DEA, focusing on applications involving the construction of composite indicators. The second objective involved a comprehensive evaluation of cities and countries aiming to promote livability and sustainable development. This evaluation was intended to contribute to the definition of better public policies through the identification of best practice examples and areas with more potential for improvements.

The first objective was addressed in chapters 3 and 4, which included the methodological developments of the thesis. Chapters 5 and 6 addressed the second objective, and corresponded to empirical applications that illustrated the applicability of the models proposed in a real-world context.

Chapter 3 approached the issue of dealing with undesirable outputs in composite indicators based on DEA models. Traditional DEA-based composite indicator models cannot be used in the presence of both desirable and undesirable outputs, as they cannot seek for reductions to the undesirable indicators. Two different approaches that can be used for the construction of CI in this context were discussed and compared. First, it was discussed an indirect approach, based on a traditional DEA model including a transformation in the measurement scale of undesirable outputs. Using a small example, it was explained that the indirect approach affects the direction of the DMUs' projection to the frontier. This occurs because the reference point used to calculate the measure of performance after transformation is no longer the origin, as in standard DEA models, but a new reference point whose coordinates depend on the constant used to transform the measurement scale of the undesirable outputs. As a consequence, this approach does not allow proportional improvements to both desirable and undesirable outputs. Furthermore, it was shown that different values for the constant used for the transformation of the outputs' measurement scale have implications both in the performance scores and ranking of the DMUs.

As an alternative to the indirect approach, it was demonstrated that a composite indicator can be derived from a DEA model specified with a directional distance function. This direct approach allows dealing with the undesirable outputs in their original measurement scale and has the advantage of preserving the proportional interpretability of the improvements. This requires setting the components of the directional vector equal to the values of the desirable and undesirable outputs of the DMU under assessment. It was concluded that this approach has better features compared to the indirect approach, and so it is more appropriate for constructing composite indicators in the presence of both desirable and undesirable outputs.

The third chapter also addressed the issue of the incorporation of restrictions to weights in the context of assessments involving composite indicators. Restrictions to weights can be included in the model in order to reflect the relative importance of individual indicators, and/or to improve the discrimination among the DMUs' performance scores. In addition, the restrictions to weights prevent DMUs from obtaining a high performance score only due to a careful choice of weights that does not reflect appropriately the true importance of the indicators. We discussed and illustrated the specification of two different types of weight restrictions that can be used in this context, namely virtual weight restrictions and assurance region type I weight restrictions. We suggested a modification to the specification of the ARI weight restrictions to allow incorporating the relative importance of outputs, expressed as a percentage. This formulation also has the advantage of being independent of the units of measurement of the output indicators and overcome some limitations of virtual weight restrictions, such as resulting in evaluations against different frontiers and the problem of having peers for inefficient DMUs that are not efficient when assessed with their own set of weights.

Chapter 4 reviewed and discussed the different approaches that can be used to accommodate undesirable outputs in the analysis of productivity change over time, namely the ratio-based Malmquist-Luenberger index and the difference-based Luenberger index. The Malmquist-Luenberger index is derived from a standard output oriented Malmquist index, using the relationship between the directional distance function and the Shephard's output distance function. It was shown that an alternative Malmquist-Luenberger index can be derived from the relationship between the directional distance function and the Shephard's input distance function. The two versions of the ML indices represent equally good adaptations of the Malmquist index for assessments involving both desirable and undesirable outputs, although they result in different values. In order to avoid the need to arbitrarily choose one of the measures, we proposed a new version of the ML index, given by the geometric mean of the two alternative versions of the ML indices. This new index, named Average ML index, has the advantage of incorporating both orientations in the computation of the productivity change score, and thus represent more accurately the changes in DMUs' features.

It was demonstrated that the Malmquist-Luenberger and Luenberger indices produce consistent and strongly correlated results. The main methodological difference between these indices that can motivate the use of one or another is related to their multiplicative or additive nature. In cases in which there is a preference for ratio-based indices, we believe that the Average Malmquist-Luenberger index proposed in this thesis should be used, as it is more precise in estimating productivity change.

This chapter also included an empirical example to compare the results obtained by the different versions of the ML indices with the Luenberger index. The results suggested that the Average ML index is more aligned with the Luenberger index than the other two versions of the ML index.

Chapter 5 used the enhanced DEA-based composite indicator constructed based on the indirect approach to treat undesirable outputs, proposed in chapter 3, to provide a single summary measure of countries environmental performance. This assessment is based on the aggregation of the indicators that underlie the estimation of the Environmental Performance Index (EPI). In addition to assigning a summary measure of performance for each country, the model applied in this chapter was used for benchmarking purposes.

This chapter also illustrated the use of the novel type of weight restrictions proposed in chapter 3, corresponding to assurance regions type I. The restrictions were used to reflect, in percentage terms, the relative importance among output indicators. Using these weight restrictions it was possible, in a first moment, to create a robust ranking of countries based on an evaluation using a common system of weights. In a second moment, by giving a range of flexibility to the weights of each indicator, it was possible to describe the strengths and weaknesses of each country and identify the peers with similar features to the under-performing countries. These features can support decision makers in the definition of better environmental policies to improve the performance of countries taking into account their specific characteristics.

Chapter 6 contributed to the development of a tool to assess and promote the livability of cities. In the first stage of this study, we proposed a conceptual model to assess cities' livability that extends the concept of urban livability. Besides including social and economic aspects, the broader concept of livability proposed in this thesis incorporates an additional component related to environmental sustainability. The resulting conceptual model is composed by 24 indicators grouped in eight dimensions, six of them related to human wellbeing, and the other two related to the environmental impact of the human living.

The second stage of the study addressed the measurement of cities' livability considering the set of indicators defined in the previous stage. The DEA-based composite indicator model used to assess cities' livability follows the direct approach to treat undesirable outputs, proposed in chapter 3, involving the use of a directional distance function.

In addition to estimating an overall measure of livability for each city, the CI allowed the specification of different directional vectors that can focus on specific components of livability (e.g., human wellbeing or environmental impact). The possibility of reflecting different priorities for improvements through the specification of different directional vectors represents an important feature of the assessment performed in this chapter.

This assessment also included the use of the ARI weight restrictions proposed in chapter 3. Two set of weight restrictions were imposed. One of them ensured that all indicators contributed to the composite indicator score, and the other ensured that the weights were homogeneously distributed among the eight dimensions of livability.

In the final stage, it was assessed the evolution of cities' performance over time. For this purpose we used the Luenberger productivity index, which is a well-established different-based index to assess productivity change over time considering simultaneous adjustments to both desirable and undesirable outputs. This analysis enabled classifying the cities in four groups, according to the profile of the changes in performance over time. The decomposition of the Luenberger productivity index allowed analysing the changes in cities' performance over time as well as to identify progression or regression movements in the production frontier. In addition, we pointed out the cities that can be considered innovators, i.e., those cities that are responsible for the shifts in the frontier towards more productive levels.

Although we illustrated in Chapter 5 and 6 that it is possible to accommodate undesirable outputs in CIs models both using a directional distance function or traditional DEA models, it is worth to reinforce the two main advantages of using the CI specified using a directional distance function. The first is related to the ability to accommodate the undesirable outputs in their original measurement scale. This feature allows to maintain the proportional interpretations of improvements to the value of the indicators, and facilitates the interpretation of the results related to the undesirable indicators. The second is related to the possibility of specifying different directional vectors. This is a very important feature that gives flexibility to customise the analysis considering different objectives or priorities for improvements in specific dimensions or indicators.

This thesis showed that DEA is a powerful technique that can be used to construct composite indicators to assess and monitor urban livability and sustainable development. One important feature of this technique is its ability to fight the subjectivity associated with the specification of the indicators' weights present in most assessments involving intangible production processes. By recurring to an optimization process, DEA allows emphasizing the best areas of each DMU under assessment and construct composite indicators with DMU-specific weights. Furthermore, DEA has the advantage of estimating performance by comparing with best practices actually observed, so it is ideal for conducting benchmarking initiatives leading to continuous improvement. Besides providing decision makers information that enables them to understand where their cities or countries stand in terms of human living conditions and sustainable development, it is also possible to identify the steps that should be taken to improve performance.

Benchmarking studies provide opportunities for cities and countries to learn with the best practices and to establish strategies to improve human living and sustainable development. Improvements in these fields can lead to competitive advantages, as investors take into account these factors when deciding where to invest in a new business or where to seek human capital.

7.2 Contributions of the thesis

This thesis contributed to the field of performance management using DEA in the presence of undesirable outputs. Innovative models were developed to address theoretical and empirical issues, as described by the following topics that summarise the main contributions of the thesis:

- The development of a DEA-based CI model, specified using a directional distance function that is able to accommodate undesirable outputs and seek for simultaneous improvements to both desirable and undesirable outputs;
- The specification of a novel type of weight restrictions, in the form of assurance region type I, that are independent of the units of measure-

ment of the outputs and allow incorporating the relative importance of outputs expressed as a percentage.

- The development of an enhanced Malmquist-Luenberger index to measure productivity change over time in assessments involving simultaneous improvements to desirable and undesirable outputs;
- The assessment of countries environmental performance, with a view to provide guidelines for performance improvements through the identification of the environmental strengths and weaknesses of each country, as well as the peers with similar features to the countries with worse performance.
- The definition of a conceptual model for the assessment of cities' livability, including components related to human wellbeing and environmental impact;
- The implementation of the CI model specified using a directional distance function to assess cities' livability, exploring different directional vectors that reflect alternative livability objectives;
- The assessment of the changes in cities' performance over time and the development of a tool to characterize the evolution of cities performance over time.

7.3 Directions for future research

The CI models developed in this thesis were applied to the evaluation of livability and sustainable development of cities and countries. These models can also be applied in different contexts that involve the production of undesirable outputs. In particular, the Directional CI would be suitable for cases where decision makers are interested in analysing DMUs' performance focusing on different objectives, which may be reflected by the specification of different directional vectors.

Concerning the empirical study of European cities' livability, the database provided by the Urban Audit program still presents a significant number of cases with missing data. As the database is being updated with the results of the 2011 census of European countries, the assessment of cities' livability could be replicated in the future when the database is more complete and robust. A larger number of cities and indicators may be included in the assessment in future analysis.

Furthermore, it would be interesting to compare the results obtained in the empirical part of this thesis, based on DEA models, with results obtained from alternative frontier methods, such as Stochastic Frontier Analysis. The comparison with other frontier methods could provide additional validation of the results.

From a methodological point of view, we believe that further research is needed on the issues related to the infeasibilities that may occur in the assessment of productivity change over time in the presence of undesirable outputs. As mentioned in this thesis, in assessments involving undesirable outputs using either the Malmquist-Luenberger or the Luenberger indices, infeasibilities occur when a DMU from one period is beyond the production possibility set of the other time period and its projection is in a direction where the frontier of the other time period does not exist. In this case it is not possible to provide an estimate of productivity change.

It would be interesting to investigate how the imposition of weight restrictions could avoid these infeasibilities. As the weight restrictions extend the segments of the frontier along rays that go to infinity, they could overcome the infeasibility problems, as the DMUs' projection to the frontier would become possible for any directional vector.

Finally, the relationship between the Average Malmquist-Luenberger and the Luenberger indices approached in this thesis could be explored in more detail. Other assessments involving larger empirical applications could provide additional validation of the results. Moreover, it would be interesting to explore the exact mathematical relationship between the Average Malmquist-Luenberger and the Luenberger indices.

Appendix A

Appendix to Chapter 4

Table A.1: Data for the 17 countries analyzed in 2005 and 2006

Country	Year	Gross output (million US$)	CO_2 emissions (thousand tons)	Capital stock (million US$)	No. of employees (thousands)	Intermediate inputs (million US$)
Austria	2005	25452241828	8209.5322	56175984363	190000	15417880136
	2006	27762073065	7647.0410	60664268509	192342	17270619633
Belgium	2005	44414843517	7500.9160	86262852958	201265	28957059955
	2006	47099955865	7418.9761	89808258206	203876	30823522302
Czech Rep.	2005	29850728941	6808.3115	56242843706	237399	18188909281
	2006	34037083557	6984.5005	62907131981	240067	19973732106
Germany	2005	206786879434	58071.2422	351617670323	1475000	123571634130
	2006	228602871410	59163.3398	381301341294	1498000	138971012688
Denmark	2005	26622581391	40325.8711	55648519338	118179	18123081233
	2006	29595043387	51251.6992	58408295881	121936	21694755675
Estonia	2005	4607849802	820.8639	4680413312	43100	3279888590
	2006	4902959852	1059.7144	4899392246	49600	3517162127
Spain	2005	98673668337	38638.2227	197595674245	672900	59050460913
	2006	109956049732	39903.4609	220501552124	710800	67187030417
Finland	2005	16728788208	4830.3315	33456437289	105900	7006965191
	2006	18179732738	4906.1582	35889864473	108300	7856612662
France	2005	138748225401	39767.0625	221997454844	1030174	72953469981
	2006	146286653584	40183.2031	236041286329	1039478	78075274585
Hungary	2005	10831614386	6701.5942	20192860021	188525	6031696414
	2006	12128323394	6863.6924	21629085077	199795	6885545213
Italy	2005	174486088251	39638.7734	322551037860	796000	105971750134
	2006	184922231887	41121.6133	346934318954	810300	114371962578
Netherlands	2005	46131523616	27915.2148	118855045021	324434	25984932109
	2006	48841180231	27294.6582	122490746641	323295	27596728198
Norway	2005	26763525209	17005.6133	75423192372	146700	17050379574
	2006	26892016526	17600.8203	76290341065	148900	16880773431
Sweden	2005	38932163916	13230.8193	56016542985	200800	25474445215
	2006	42194486689	13751.1621	60133980308	200100	27901427938
Slovenia	2005	3978262193	4397.8052	7963174037	35105	2474600861
	2006	4484790406	4615.687	8872307623	36091	2827293529
Slovakia	2005	9443120615	1842.4961	17016397167	97097	5760523824
	2006	9467091320	1707.2931	18220063583	100932	6216120302
U. Kingdom	2005	170422611254	99081.1484	212578331137	1096426	101019081095
	2006	177010969354	92436.0078	227010338818	1094101	106130293799

Appendix B

Appendix to Chapter 5, part 1

The formulation below corresponds to the dual of the model presented in (3.2) with the restrictions shown in (5.2). It was used to address the third objective of this chapter, that is related with the identification of the peers that countries with worse performance should look in order to search for examples of best practices. This formulation is known as the envelopment formulation of a DEA model. The variables λ_j $(j = 1, ..., n)$ allow to identify the peers (i.e. benchmarks) for the inefficient DMUs. The peers for the DMU j_0 under assessment are the units that present values of λ_j^* greater than zero at the optimal solution of model (B.1). The variables α_r, β_r, ϑ_k and φ_k are the dual variables associated to the weight restrictions shown in (5.2).

$$
\begin{aligned}
&\text{Max } \theta && \text{(B.1)}\\
&y_{rj_0}\ \theta - \sum_{j=1}^{n} y_{rj}\ \lambda_j\ + \\
&\quad + \bar{y}_r\ \alpha_r - \bar{y}_r\ \left[\sum_{r'=1}^{s} w_{r'}(1-c)\ \alpha_{r'} + \sum_{k'=1}^{l} w_{k'}(1-c)\ \vartheta_{k'}\right] - \\
&\quad - \bar{y}_r\ \beta_r + \bar{y}_r \left[\sum_{r'=1}^{s}\ w_{r'}(1+c)\ \beta_{r'} + \sum_{k'=1}^{l} w_{k'}(1-c)\ \varphi_{k'}\right] \leq 0 \quad r = 1, \ldots, s \\
&(M_k - b_{kj_0})\ \theta - \sum_{j=1}^{n} (M_k - b_{kj})\lambda_j\ + \\
&\quad + (M_k - \bar{b}_k)\ \vartheta_k - (M_k - \bar{b}_k) \left[\sum_{k'=1}^{l} w_{k'}(1-c)\ \vartheta_{k'} + \sum_{r'=1}^{s} w_{r'}(1-c)\ \alpha_{r'}\right] - \\
&\quad - (M_k - \bar{b}_k)\ \varphi_k + (M_k - \bar{b}_k) \left[\sum_{k'=1}^{l} w_{k'}(1-c)\ \varphi_{k'} + \sum_{r'=1}^{s} w_{r'}(1-c)\ \beta_{r'}\right] \leq 0 \\
& && k = 1, \ldots, l \\
&\sum_{j=1}^{n} \lambda_j \leq 1 && j = 1, \ldots, n \\
&\lambda_j, \alpha_r, \beta_r, \vartheta_k\ and\ \varphi_k \geq 0 && \forall\ j,\ r\ and\ k \\
&\delta\ isfree
\end{aligned}
$$

The objective function value at the optimal solution corresponds to the factor θ^* by which all outputs of the DMU under assessment can be proportionally increased to reach the target output values. The efficiency score of DMU j_0 under assessment is the reciprocal of this value. Therefore, the DMUs considered efficient are those for which there is no evidence that it is possible to expand their outputs, such that the value of θ^* is equal to 1.

Appendix C

Appendix to Chapter 5, part 2

Table C.1: Results obtained in each cluster and comparison between the DEA scores and EPI scores

	Country	DEA Score (rank)	EPI Score (rank)	Country	DEA Score (rank)	EPI Score (rank)
C1	Nicaragua	100% (1)	57.1% (93)	Nigeria	63.4% (19)	40.2% (153)
	Kenya	85.1% (2)	51.4% (108)	Benin	62.3% (20)	39.6% (154)
	Cameroon	84.3% (3)	44.6% (133)	Zambia	61.8% (21)	47.0% (130)
	Swaziland	83.1% (4)	54.4% (101)	Guinea	60.2% (22)	44.4% (136)
	Eritrea	82.7% (5)	54.6% (100)	Malawi	60.0% (23)	51.4% (107)
	Nepal	78.8% (6)	68.2% (38)	Uganda	59.7% (24)	49.8% (119)
	Congo	77.8% (7)	54.0% (105)	Togo	59.0% (25)	36.4% (159)
	Mauritania	77.1% (8)	33.7% (161)	E. Guinea	58.4% (26)	41.9% (146)
	Zimbabwe	76.3% (9)	47.8% (127)	Ethiopia	52.6% (27)	43.1% (141)
	Ghana	75.6% (10)	51.3% (109)	C. African Rep.	51.8% (28)	33.3% (162)
	P. N. Guinea	72.9% (11)	44.3% (138)	Burundi	51.7% (29)	43.9% (140)
	Laos	72.9% (12)	59.6% (80)	Rwanda	50.3% (30)	44.6% (135)
	Cte d'Ivoire	70.8% (13)	54.3% (102)	Guinea-Bissau	50.0% (31)	44.7% (132)
	Tanzania	67.5% (14)	47.9% (126)	Burkina Faso	46.7% (32)	47.3% (128)
	Mozambique	65.7% (15)	51.2% (112)	Chad	43.9% (33)	40.8% (151)
	Cambodia	65.2% (16)	41.7% (148)	Mali	43.3% (34)	39.4% (156)
	Madagascar	65.0% (17)	49.2% (120)	Sierra Leone	40.9% (35)	32.1% (163)
	Haiti	64.4% (18)	39.5% (155)	Niger	36.8% (36)	37.6% (158)
	Spearman correlation: 0.6302 ($p = 0.000019$)					
C2	Costa Rica	100% (1)	86.4% (3)	Romania	90.3% (23)	67.0% (45)
	Ecuador	98.8% (2)	69.3% (30)	Canada	89.8% (24)	66.4% (46)
	Iceland	98.7% (3)	93.5% (1)	Croatia	89.7% (25)	68.7% (35)
	France	97.9% (4)	78.2% (7)	Ireland	89.4% (26)	67.1% (44)
	Colombia	97.6% (5)	76.8% (10)	Austria	89.3% (27)	78.1% (8)
	Germany	97.4% (6)	73.2% (17)	Malaysia	89.0% (28)	65.0% (54)
	Portugal	97.3% (7)	73.0% (19)	Russia	88.9% (29)	61.2% (69)
	Sweden	96.0% (8)	86.0% (4)	Netherlands	88.0% (30)	66.4% (47)
	Italy	95.8% (9)	73.1% (18)	Slovakia	87.2% (31)	74.5% (13)
	Dominican Rep.	95.7% (10)	68.4% (36)	Hungary	86.9% (32)	69.1% (33)
	United Kingdom	94.7% (11)	74.2% (14)	Bulgaria	86.9% (33)	62.5% (65)
	Spain	94.7% (12)	70.6% (25)	Poland	86.8% (34)	63.1% (63)
	Cuba	94.5% (13)	78.1% (9)	Estonia	86.8% (35)	63.8% (57)
	New Zealand	93.4% (14)	73.4% (15)	Slovenia	86.6% (36)	65.0% (55)
	Panama	93.0% (15)	71.4% (24)	Greece	86.4% (37)	60.9% (71)
	Norway	92.9% (16)	81.1% (5)	Czech Rep.	85.9% (38)	71.6% (22)

Continued on next page

Table C.1 – *Continued from previous page*

	Country	DEA Score (rank)	EPI Score (rank)	Country	DEA Score (rank)	EPI Score (rank)
	Japan	92.7% (17)	72.5% (20)	Malta	85.5% (39)	76.3% (11)
	Switzerland	91.7% (18)	89.1% (2)	South Korea	84.9% (40)	57.0% (94)
	Latvia	91.5% (19)	72.5% (21)	Belgium	83.9% (41)	58.1% (88)
	Finland	91.2% (20)	74.7% (12)	Luxembourg	81.5% (42)	67.8% (41)
	Lithuania	90.6% (21)	68.3% (37)	Cyprus	78.4% (43)	56.3% (96)
	Denmark	90.5% (22)	69.2% (32)			
	Spearman correlation: 0.6855 ($p = 0.0000002$)					
C3	Mauritius	100% (1)	80.6% (6)	Morocco	85.7% (24)	65.6% (52)
	Chile	96.8% (2)	73.3% (16)	Lebanon	85.4% (25)	57.9% (90)
	Australia	95.5% (3)	65.7% (51)	S.& Montenegro	84.7% (26)	69.4% (29)
	Venezuela	94.3% (4)	62.9% (64)	Suriname	84.3% (27)	68.2% (39)
	T. & Tobago	93.3% (5)	54.2% (103)	Tunisia	83.9% (28)	60.6% (74)
	Mexico	93.1% (6)	67.3% (43)	Gabon	82.4% (29)	56.4% (95)
	Jordan	93.0% (7)	56.1% (97)	Maldives	82.1% (30)	65.9% (48)
	Israel	92.9% (8)	62.4% (66)	Kyrgyzstan	81.4% (31)	59.7% (79)
	Belize	92.9% (9)	69.9% (26)	Guyana	80.9% (32)	59.2% (82)
	El Salvador	91.4% (10)	69.1% (34)	Azerbaijan	80.8% (33)	59.1% (84)
	Algeria	90.5% (11)	67.4% (42)	Kazakhstan	80.6% (34)	57.3% (92)
	Brunei D.	89.8% (12)	60.8% (72)	Macedonia	79.7% (35)	60.6% (73)
	Iran	89.5% (13)	60.0% (78)	Djibouti	79.7% (36)	60.5% (75)
	South Africa	89.1% (14)	50.8% (115)	B.& Herzegovina	79.6% (37)	55.9% (98)
	Ukraine	89.0% (15)	58.2% (87)	Moldova	78.4% (38)	58.8% (86)
	Turkey	88.3% (16)	60.4% (77)	Oman	75.6% (39)	45.9% (131)
	A. & Barbuda	88.0% (17)	69.8% (27)	Uzbekistan	74.0% (40)	42.3% (144)
	Egypt	88.0% (18)	62.0% (68)	Yemen	71.7% (41)	48.3% (124)
	Syria	87.4% (19)	64.6% (56)	S.T. & Principe	71.4% (42)	57.3% (91)
	Jamaica	87.2% (20)	58.0% (89)	Solomon Islands	71.4% (43)	51.1% (114)
	Belarus	86.3% (21)	65.4% (53)	Turkmenistan	68.5%(44)	38.4% (157)
	Georgia	85.9% (22)	63.6% (59)	Iraq	66.1% (45)	41.0% (150)
	Armenia	85.8% (23)	60.4% (76)	Botswana	64.5% (46)	41.3% (149)
	Spearman correlation: 0.6252 ($p = 0.000002$)					
C4	Peru	100% (1)	69.3% (31)	China	85.3% (15)	49.0% (121)
	Thailand	99.9% (2)	62.2% (67)	Pakistan	82.0% (16)	48.0% (125)
	Albania	99.7% (3)	71.4% (23)	Tajikistan	81.9% (17)	51.3% (111)
	Argentina	99.3% (4)	61.0% (70)	Myanmar	80.9% (18)	51.3% (110)
	Philippines	98.3% (5)	65.7% (50)	India	79.2% (19)	48.3% (123)
	Fiji	95.5% (6)	65.9% (49)	Bhutan	78.6% (20)	68.0% (40)
	Sri Lanka	92.9% (7)	63.7% (58)	Bolivia	76.8% (21)	44.3% (137)
	Uruguay	92.9% (8)	59.1% (83)	Gambia	75.2% (22)	50.3% (116)
	Honduras	90.6% (9)	49.9% (118)	North Korea	75.1% (23)	41.8% (147)
	Viet Nam	89.5% (10)	59.0% (85)	Sudan	71.8% (24)	47.1% (129)
	Namibia	89.2% (11)	59.3% (81)	Mongolia	71.1% (25)	42.8% (142)
	Guatemala	89.0% (12)	54.0% (104)	Senegal	68.4% (26)	42.3% (143)
	Paraguay	86.6% (13)	63.5% (60)	Bangladesh	68.1% (27)	44.0% (139)
	Indonesia	85.6% (14)	44.6% (134)			
	Spearman correlation: 0.7982 ($p = 0.0000003$)					

Appendix D

Appendix to Chapter 6, part 1

Formulation (D.1) corresponds to the dual of the model presented in (3.8) with the restrictions shown in (6.1) and (6.2). It was used to identify the peers that cities with worse performance should look in order to search for examples of best practices. The peers for the DMU j_0 under assessment are the units that present values of λ_j^* greater than zero at the optimal solution of model (D.1). The variables α_r, γ_k, ϑ_z and φ_z are the dual variables associated to the weight restrictions shown in (6.1) and (6.2).

$$
\begin{aligned}
&\text{Max } \beta && \text{(D.1)}\\
&g_{y_{rj_0}}\ \beta - \sum_{j=1}^{n} y_{rj}\ \lambda_j\ + \bar{y}_r\ \alpha_r - 0.01\ \bar{y}_r \left[\sum_{r'=1}^{s}\ \alpha_{r'} + \sum_{k'=1}^{l}\ \gamma_{k'}\right] + \\
&\qquad + \bar{y}_r\ \vartheta_{z_{(r)}} - 0.10\ \bar{y}_r\ \sum_{z=1}^{q}\ \vartheta_z\ - \\
&\qquad - \bar{y}_r\ \varphi_{z_{(r)}} + 0.15\ \bar{y}_r\ \sum_{z=1}^{q}\ \varphi_z \leq - \sum_{j=1}^{n} y_{rj_0} && r = 1, \ldots, s\\
&g_{b_{kj_0}}\ \beta + \sum_{j=1}^{n} b_{kj}\ \lambda_j\ + \bar{b}_k\ \gamma_k - 0.01\ \bar{b}_k \left[\sum_{r'=1}^{s}\ \alpha_{r'} + \sum_{k'=1}^{l}\ \gamma_{k'}\right] + \\
&\qquad + \bar{b}_k\ \vartheta_{z_{(k)}} - 0.10\ \bar{b}_k\ \sum_{z=1}^{q}\ \vartheta_z\ - \\
&\qquad - \bar{b}_k\ \varphi_{z_{(k)}} + 0.15\ \bar{b}_k\ \sum_{z=1}^{q}\ \varphi_z = \sum_{j=1}^{n} b_{kj_0} && k = 1, \ldots, l\\
&\sum_{j=1}^{n} \lambda_j \leq 1 && j = 1, \ldots, n\\
&\lambda_j, \alpha_r, \gamma_k, \vartheta_z, \varphi_z \geq 0 && \forall\ j,\ r,\ k\ and\ z\\
&\beta\ isfree
\end{aligned}
$$

The indices $z_{(r)}$ and $z_{(k)}$ refer to the dimension z for which the outputs r and k belong. Note that the indices $z_{(r)}$ and $z_{(k)}$ are always less than or equal to q, where q is the total number of dimensions.

The objective function value at the optimal solution of the model (D.1) corresponds to the maximal feasible expansion of desirable outputs and contraction of undesirable outputs that can be achieved simultaneously.

Appendix E

Appendix to Chapter 6, part 2

Table E.1: Results for all cities assessed

Country	City	Classification	Luenberger Index	Efficiency Change	Technical Change	Livability 2003-2006	Livability 2007-2009	EI 2007-2009	HWB 2007-2009
Spain	S. de Compostela*	star	0.108	0	0.108	0	0	0	0
Bulgaria	Plovdiv	star	0.101	0.043	0.058	0.068	0.025	0.140$^{\blacktriangle}$	0.031
Portugal	Funchal*	star	0.090	0	0.090	0	0	0	0
Bulgaria	Ruse*	star	0.071	0	0.071	0	0	0	0
Bulgaria	Sofia*	star	0.066	0.047	0.019	0.047	0	0	0
Slovakia	Trencín*	star	0.060	0	0.060	0	0	0	0
Estonia	Tallinn*	star	0.059	0.028	0.031	0.028	0	0	0
Norge	Tromso*	star	0.059	0	0.059	0	0	0	0
Belgium	Brugge*	star	0.050	0	0.050	0	0	0	0
Slovakia	PreSov*	star	0.050	0	0.050	0	0	0	0
Norge	Bergen*	star	0.049	0.013	0.036	0.013	0	0	0
Spain	Pamplona/Iruña*	star	0.046	0	0.046	0	0	0	0
Switzerland	Bern*	star	0.046	0	0.046	0	0	0	0
Slovakia	Banska Bystrica*	star	0.040	0	0.040	0	0	0	0
Sweden	Goteborg*	star	0.036	0.033	0.003	0.033	0	0	0
Poland	Rzeszow	star	0.035	0.016	0.019	0.034	0.018	0.093$^{\blacktriangle}$	0.023$^{\blacktriangle}$
Germany	Potsdam*	star	0.033	0.023	0.010	0.023	0	0	0
Germany	Mainz*	star	0.032	0	0.032	0	0	0	0
Bulgaria	Burgas*	star	0.032	0	0.032	0	0	0	0
Germany	Weimar*	star	0.030	0	0.030	0	0	0	0
Slovakia	Zilina	star	0.026	-0.003	0.029	0.030	0.033	0.160	0.040
Portugal	Lisbon*	star	0.018	0	0.018	0	0	0	0
Germany	Munich	star	0.014	0.033	-0.019	0.050	0.017	0.175$^{\blacktriangle}$	0.019
Germany	Frankfurt Main*	star	0.013	0	0.013	0	0	0	0
Belgium	Brussels	star	0.011	0.017	-0.005	0.051	0.034	0.282$^{\blacktriangle}$	0.038
Belgium	Liege	star	0.011	-0.006	0.018	0.064	0.071	0.437$^{\blacktriangle}$	0.079
Spain	Valencia	star	0.010	-0.021	0.031	0.031	0.051	0.413$^{\blacktriangle}$	0.058
Germany	Bochum*	star	0.008	0	0.008	0	0	0	0
Spain	Toledo	star	0.008	-0.067	0.0750	0	0.067	0.320	0.079
Germany	Darmstadt*	star	0.005	0	0.005	0	0	0	0
Germany	Bielefeld	star	0.002	0.020	-0.018	0.056	0.036	0.372$^{\blacktriangle}$	0.039
Finland	Tampere*	star	0.001	0	0.001	0	0	0	0
Spain	Las Palmas	rising star	0.150	0.117	0.033	0.229	0.111	0.487	0.136
Poland	Kielce	rising star	0.136	0.125	0.012	0.300	0.176	0.706	0.232$^{\blacktriangle}$
Latvia	Riga	rising star	0.126	0.103	0.024	0.294	0.191	0.670	0.263$^{\blacktriangle}$
Slovenia	Ljubljana	rising star	0.125	0.105	0.019	0.124	0.018	0.100$^{\blacktriangle}$	0.022
Slovenia	Maribor	rising star	0.116	0.117	-0.001	0.264	0.147	0.584	0.192$^{\blacktriangle}$
Slovakia	Nitra	rising star	0.108	0.05	0.058	0.143	0.092	0.381	0.118$^{\blacktriangle}$

Continued on next page

Table E.1 – *Continued from previous page*

Country	City	Classification	Luenberger Index	Efficiency Change	Technical Change	Livability 2003-2006	Livability 2007-2009	EQ	HWB
Lithuania	Panevezys	rising star	0.107	0.092	0.015	0.132	0.041	0.240$^{\blacktriangle}$	0.049
Poland	Konin*	rising star	0.103	0.088	0.015	0.088	0	0	0
Germany	Dresden	rising star	0.098	0.108	-0.011	0.117	0.008	0.051$^{\blacktriangle}$	0.01
Spain	Palma de Mallorca	rising star	0.096	0.076	0.020	0.127	0.052	0.308$^{\blacktriangle}$	0.061
Bulgaria	Varna	rising star	0.086	0.048	0.038	0.166	0.118	0.515	0.154$^{\blacktriangle}$
Poland	Krakow	rising star	0.078	0.069	0.009	0.219	0.149	0.695	0.183
Poland	Opole	rising star	0.074	0.064	0.010	0.181	0.117	0.489	0.154$^{\blacktriangle}$
Spain	Cordoba	rising star	0.073	0.037	0.036	0.187	0.150	0.618	0.187$^{\blacktriangle}$
Germany	Koblenz	rising star	0.071	0.076	-0.005	0.076	0	0	0
Poland	Nowy Sacz	rising star	0.071	0.061	0.010	0.125	0.064	0.290	0.081$^{\blacktriangle}$
Bulgaria	Vidin	rising star	0.067	0.001	0.066	0.100	0.098	0.440	0.124$^{\blacktriangle}$
Finland	Oulu	rising star	0.064	0.074	-0.010	0.074	0	0	0
Spain	Madrid	rising star	0.058	0.054	0.004	0.080	0.026	0.174$^{\blacktriangle}$	0.031
Poland	Olsztyn	rising star	0.056	0.051	0.005	0.135	0.084	0.421$^{\blacktriangle}$	0.105$^{\blacktriangle}$
Poland	Torun	rising star	0.056	0.053	0.003	0.193	0.140	0.688	0.176$^{\blacktriangle}$
Poland	Gorzow Wlkp	rising star	0.055	0.042	0.013	0.208	0.166	0.728	0.214$^{\blacktriangle}$
Lithuania	Vilnius	rising star	0.055	0.030	0.025	0.073	0.043	0.255$^{\blacktriangle}$	0.051
Slovakia	Bratislava	rising star	0.053	0.030	0.023	0.114	0.084	0.359	0.110$^{\blacktriangle}$
Poland	Poznan	rising star	0.051	0.033	0.018	0.19	0.156	0.631	0.205$^{\blacktriangle}$
Poland	Warszawa	rising star	0.049	0.025	0.024	0.164	0.139	0.591	0.175$^{\blacktriangle}$
Poland	Katowice	rising star	0.048	0.044	0.004	0.221	0.177	0.732	0.233$^{\blacktriangle}$
Portugal	Porto	rising star	0.044	0.03	0.014	0.120	0.09	0.465$^{\blacktriangle}$	0.107
Slovakia	Trnava	rising star	0.040	0.014	0.026	0.169	0.155	0.614	0.200$^{\blacktriangle}$
Germany	Dortmund	rising star	0.036	0.044	-0.009	0.184	0.140	0.698$^{\blacktriangle}$	0.166
Poland	Zielona Gora	rising star	0.035	0.035	0	0.163	0.129	0.62	0.162$^{\blacktriangle}$
Lithuania	Kaunas	rising star	0.035	-0.007	0.042	0.197	0.204	0.743	0.269$^{\blacktriangle}$
Germany	Leipzig	rising star	0.031	0.035	-0.004	0.17	0.135	0.682$^{\blacktriangle}$	0.164
Germany	Wiesbaden	rising star	0.027	0.014	0.014	0.104	0.091	0.488$^{\blacktriangle}$	0.111
Portugal	Braga	rising star	0.026	0.045	-0.02	0.096	0.05	0.274$^{\blacktriangle}$	0.061
Germany	Hannover	rising star	0.025	0.017	0.008	0.078	0.061	0.337$^{\blacktriangle}$	0.074
Bulgaria	Pleven	rising star	0.022	-0.042	0.063	0.071	0.113	0.471	0.143$^{\blacktriangle}$
Germany	Saarbrucken	rising star	0.020	0.06	-0.040	0.162	0.102	0.714$^{\blacktriangle}$	0.114
Germany	Mulheim a.d.Ruhr	rising star	0.018	0.023	-0.005	0.104	0.082	0.447$^{\blacktriangle}$	0.095
Spain	Zaragoza	rising star	0.015	0.026	-0.011	0.15	0.124	0.531	0.155$^{\blacktriangle}$
Germany	Monchengladbach	rising star	0.015	0.041	-0.026	0.119	0.079	0.528$^{\blacktriangle}$	0.092
Estonia	Tartu	rising star	0.014	-0.001	0.015	0.120	0.121	0.495	0.160$^{\blacktriangle}$
Spain	Malaga	rising star	0.014	0	0.014	0.154	0.154	0.598	0.187
Germany	Stuttgart	rising star	0.010	0.035	-0.025	0.091	0.056	0.324$^{\blacktriangle}$	0.066
Germany	Augsburg	rising star	0.007	0.013	-0.006	0.088	0.076	0.438$^{\blacktriangle}$	0.089

Continued on next page

Table E.1 – *Continued from previous page*

Country	City	Classification	Luenberger Index	Efficiency Change	Technical Change	Livability 2003-2006	Livability 2007-2009	EQ	HWB
Germany	Berlin	rising star	0.005	0.032	-0.027	0.118	0.086	0.483$^{\blacktriangle}$	0.100
Germany	Kiel	rising star	0.002	0.030	-0.028	0.135	0.105	0.638$^{\blacktriangle}$	0.124
Finland	Turku	rising star	0.002	-0.003	0.005	0.086	0.090	0.393	0.113$^{\blacktriangle}$
Germany	Dusseldorf	falling star	-0.001	0.005	-0.006	0.006	0	0	0
Slovakia	Kosice	falling star	-0.005	-0.023	0.018	0.068	0.091	0.41	0.117$^{\blacktriangle}$
Spain	Santander	falling star	-0.006	-0.029	0.024	0	0.029	0.240$^{\blacktriangle}$	0.033
Germany	Bremen	falling star	-0.007	0.010	-0.017	0.067	0.057	0.379$^{\blacktriangle}$	0.065
Switzerland	Zurich	falling star	-0.008	0	-0.008	0	0	0	0
Germany	Schwerin	falling star	-0.009	0.015	-0.024	0.015	0	0	0
Germany	Nurnberg	falling star	-0.014	0.013	-0.027	0.013	0	0	0
Belgium	Gent	falling star	-0.014	-0.026	0.012	0.024	0.051	0.307$^{\blacktriangle}$	0.059
Hungary	Budapest	falling star	-0.018	-0.039	0.020	0	0.039	0.191	0.048$^{\blacktriangle}$
Belgium	Charleroi	falling star	-0.019	-0.021	0.001	0.055	0.076	0.472$^{\blacktriangle}$	0.089
Germany	Cologne	falling star	-0.022	0.012	-0.034	0.048	0.036	0.315$^{\blacktriangle}$	0.040
Germany	Regensburg	falling star	-0.022	0	-0.022	0	0	0	0
Germany	Bonn	falling star	-0.023	0	-0.023	0	0	0	0
Germany	Karlsruhe	falling star	-0.023	-0.023	0	0	0.023	0.153$^{\blacktriangle}$	0.027
Germany	Hamburg	falling star	-0.027	-0.002	-0.025	0.063	0.065	0.478$^{\blacktriangle}$	0.071
Norge	Stavanger	falling star	-0.028	0	-0.028	0	0	0	0
Spain	Barcelona	falling star	-0.034	-0.005	-0.029	0	0.005	0.030$^{\blacktriangle}$	0.006
Sweden	Stockholm	falling star	-0.036	0	-0.036	0	0	0	0
Spain	Valladolid	falling star	-0.038	-0.080	0.042	0	0.08	0.402$^{\blacktriangle}$	0.093
Spain	Oviedo	falling star	-0.042	-0.078	0.036	0.001	0.079	0.398$^{\blacktriangle}$	0.095
Germany	Trier	falling star	-0.044	-0.034	-0.009	0.045	0.079	0.376	0.097
Germany	Freiburg Breisgau	falling star	-0.054	-0.047	-0.007	0	0.047	0.250$^{\blacktriangle}$	0.056
Hungary	Pecs	falling star	-0.057	-0.039	-0.018	0	0.039	0.187	0.049$^{\blacktriangle}$
Belgium	Antwerpen	falling star	-0.060	-0.017	-0.043	0	0.017	0.169$^{\blacktriangle}$	0.019
Germany	Gottingen	falling star	-0.065	-0.057	-0.008	0	0.057	0.300$^{\blacktriangle}$	0.070
Austria	Wien	falling star	-0.068	-0.009	-0.060	0	0.009	0.071$^{\blacktriangle}$	0.010
Spain	Murcia	falling star	-0.171	-0.127	-0.044	0	0.127	0.631$^{\blacktriangle}$	0.151
Spain	Badajoz	falling star	-0.198	-0.133	-0.066	0	0.133	0.709$^{\blacktriangle}$	0.163
Germany	Essen	problem	-0.001	0.017	-0.019	0.072	0.055	0.433$^{\blacktriangle}$	0.060
Spain	Sevilla	problem	-0.002	-0.013	0.011	0.188	0.201	0.745	0.257$^{\blacktriangle}$
Spain	Logrono	problem	-0.003	-0.036	0.033	0.100	0.136	0.581	0.169$^{\blacktriangle}$
Germany	Halle an der Saale	problem	-0.010	-0.024	0.014	0.100	0.124	0.614$^{\blacktriangle}$	0.151
Germany	Frankfurt (Oder)	problem	-0.011	-0.009	-0.002	0.080	0.089	0.535$^{\blacktriangle}$	0.107
Germany	Erfurt	problem	-0.013	0.015	-0.028	0.164	0.149	0.913$^{\blacktriangle}$	0.168
Spain	Vitoria/Gasteiz	problem	-0.016	0.004	-0.020	0.081	0.077	0.398$^{\blacktriangle}$	0.090
Germany	Moers	problem	-0.019	-0.002	-0.017	0.130	0.133	0.755$^{\blacktriangle}$	0.147

Continued on next page

Table E.1 – *Continued from previous page*

Country	City	Classification	Luenberger Index	Efficiency Change	Technical Change	Livability 2003-2006	Livability 2007-2009	EQ	HWB
UK	Portsmouth	problem	-0.027	-0.03	0.004	0.118	0.148	0.653	0.180
Portugal	Setubal	problem	-0.034	-0.021	-0.013	0.197	0.218	0.863	0.272▲
Sweden	Malmo	problem	-0.048	-0.015	-0.033	0.095	0.110	0.690▲	0.122
Germany	Magdeburg	problem	-0.072	-0.045	-0.027	0.111	0.156	0.767	0.189
Average scores			0.021	0.012	0.009	0.078	0.066	0.325	0.081

* innovative cities

▲ Potential for improvement above the expected value in the specific component

References

Aigner, D. J. and Chu, S. F. (1968). On estimating the industry production function. *The American Economic Review*, 58(4):826–839.

Aigner, D. J., Lovell, C. A. K., and Schmidt, P. (1977). Formulation and estimation of stochastic frontier production function models. *Journal of Econometrics*, 6(1):21–37.

Ali, A. I. and Seiford, L. M. (1990). Translation-invariance in data envelopment analysis. *Operations research letters*, 9(6):403–405.

Allen, R., Athanassopoulos, A. D., Dyson, R. G., and Thanassoulis, E. (1997). Weights restrictions and value judgments in data envelopment analysis: evolution, development and future directions. *Annals of Operations Research*, 0(73):13–64.

Angulo-Meza, L. and Lins, M. E. (2002). Review of methods for increasing discrimination in data envelopment analysis. *Annals of Operations Research*, 116(1-4):225–242.

Aparicio, J., Pastor, J. T., and Zofio, J. L. (2013). On the inconsistency of the MalmquistLuenberger index. *European Journal of Operational Research*, 229(3):738–742.

Australian Department of Infrastructure and Transport (2010). State of australian cities 2010. Technical report, Department of Infrastructure and Transport - Major Cities Unit, Canberra, Australia.

Australian Department of Infrastructure and Transport (2012). State of australian cities 2012. Technical report, Department of Infrastructure and Transport - Major Cities Unit, Canberra, Australia.

Banker, R. D. (1984). Estimating most productive scale size using data envelopment analysis. *European Journal of Operational Research*, 17(1):35–44.

Banker, R. D. (1993). Maximum likelihood, consistency and data envelopment analysis: A statistical foundation. *Management Science*, 39(10):1265–1273.

Banker, R. D. (1996). Hypothesis tests using data envelopment analysis. *Journal of Productivity Analysis*, 7(2-3):139–159.

Banker, R. D., Charnes, A., and Cooper, W. W. (1984). Some models for estimating technical and scale inefficiencies in data envelopment analysis. *Management Science*, 30(9):1078–1092.

Banker, R. D. and Thrall, R. M. (1992). Estimation of returns to scale using data envelopment analysis. *European Journal of Operational Research*, 62(1):74–84.

Boussemart, J. P., Briec, W., Kerstens, K., and Poutineau, J.-C. (2003). Luenberger and malmquist productivity indices: Theoretical comparisons and empirical illustration. In *Bulletin of Economic Research, Bulletin of Economic Research*, volume 55, pages 391–405.

Briec, W., Peypoch, N., and Ratsimbanierana, H. (2011). Productivity growth and biased technological change in hydroelectric dams. *Energy Economics*, 33(5):853–858.

Burck, J., Bals, C., and Rossow, V. (2009). Climate change performance index. Technical report, Germanwatch and Climate Action Network Europe, Berlin, Germany.

Burck, J., Hermwille, L., and Krings, L. (2012). Climate change performance index. Technical report, Germanwatch and Climate Action Network Europe, Berlin, Germany.

Caves, D. W., Christensen, L. R., and Diewert, E. (1982). Multilateral comparisons of output, input, and productivity using superlative index numbers. *The Economic Journal*, 92(365):73–86.

Chambers, R. G. (1996). A new look at exact input, output, and productivity measurement. *Maryland Agricultural Experiment Station.*

Chambers, R. G., Chung, Y., and Fare, R. (1996). Benefit and distance functions. *Journal of Economic Theory*, 70(2):407–419.

Charnes, A., Cooper, W. W., and Rhodes, E. (1978). Measuring the efficiency of decision making units. *European Journal of Operational Research*, 2(6):429–444.

Cherchye, L., Moesen, W., Rogge, N., and Van Puyenbroeck, T. (2007). An introduction to Benefit of the doubt composite indicators. *Social Indicators Research*, 82(1):111–145.

Cherchye, L., Moesen, W., Rogge, N., and Van Puyenbroeck, T. (2011). Constructing composite indicators with imprecise data: A proposal. *Expert Systems with Applications*, 38(9):10940–10949.

Cherchye, L., Moesen, W., Rogge, N., Van Puyenbroeck, T., Saisana, M., Saltelli, A., Liska, R., and Tarantola, S. (2008). Creating composite indicators with DEA and robustness analysis: the case of the technology achievement index. *Journal of the Operational Research Society*, 59(2):239–251.

Chung, Y., Fare, R., and Grosskopf, S. (1997). Productivity and undesirable outputs: A directional distance function approach. *Journal of Environmental Management*, 51(3):229–240.

Coelli, T. J., Rao, D. S. P., ODonnell, C. J., and Battese, G. E. (2005). *An Introduction to Efficiency and Productivity Analysis.* Springer Science+Business Media, Inc., New York, United States, 2. edition.

Cook, W. D. and Green, R. H. (2005). Evaluating power plant efficiency: a hierarchical model. *Computers & Operations Research*, 32(4):813–823.

Cook, W. D. and Kress, M. (1990). A data envelopment model for aggregating preference rankings. *Management Science*, 36(11):1302–1310.

Davies, D. and Bouldin, D. (1979). A cluster separation measure. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 1:224–227.

Debreu, G. (1951). The coefficient of resource utilization. *Econometrica*, 19(3):273–292.

Deprins, D. and Tulkens, H. (1984). Measuring labor-efficiency in post offices. In *The Performance of Public Enterprises: Concepts and Measurement*, pages 243–267. Elsevier, Amsterdam.

Despotis, D. K. (2002). Improving the discriminating power of DEA: focus on globally efficient units. *Journal of the Operational Research Society*, 53(3):314–323.

Despotis, D. K. (2004). A reassessment of the human development index via data envelopment analysis. *Journal of the Operational Research Society*, 56(8):969–980.

Despotis, D. K. (2005). Measuring human development via data envelopment analysis: the case of asia and the pacific. *Omega*, 33(5):385–390.

Dyson, R. G., Allen, R., Camanho, A. S., Podinovski, V. V., Sarrico, C. S., and Shale, E. A. (2001). Pitfalls and protocols in DEA. *European Journal of Operational Research*, 132(2):245–259.

Dyson, R. G. and Thanassoulis, E. (1988). Reducing weight flexibility in data envelopment analysis. *Journal of the Operational Research Society*, 39(6):563–576.

Emerson, J. W., Esty, D. C., Levy, M. A., Kim, C. H., Mara, V., Sherbinin, A., and Srebotnjak, T. (2010). 2010 environmental performance index. Technical report, Yale Center for Environmental Law and Policy, New Haven.

Emerson, J. W., Hsu, A., Levy, M. A., de Sherbinin, A., Mara, V., Esty, D. C., and Jaiteh, M. (2012). Environmental performance index and pilot trend environmental performance index. Technical report, Yale Center for Environmental Law and Policy, New Haven.

English Department for Communities and Local Government (2006). State of the english cities: Liveability in english cities. Technical report, Department for Communities and Local Government, London, England.

Epure, M., Kerstens, K., and Prior, D. (2011). Bank productivity and performance groups: A decomposition approach based upon the luenberger productivity indicator. *European Journal of Operational Research*, 211(3):630–641.

European Union (2008). Energy and climate change elements of the final compromise. Technical report, European Union.

Fare, R. and Grosskopf, S. (2000). Theory and application of directional distance functions. *Journal of Productivity Analysis*, 13(2):93–103.

Fare, R. and Grosskopf, S. (2005). *New directions: efficiency and productivity*. Springer Science+Business Media, Inc., New York, 1 edition.

Fare, R., Grosskopf, S., Lovell, C., and Yaisawarng, S. (1993). Derivation of shadow prices for undesirable outputs: A distance function approach. *The review of economics and statistics*, 75(2):374–380.

Fare, R., Grosskopf, S., and Lovell, C. A. K. (1985). *The Measurement of Efficiency of Production*. Klumer-Nijhoff Publishing, Boston.

Fare, R., Grosskopf, S., and Lovell, C. A. K. (1994a). *Production Frontiers*. Cambridge University Press, Cambridge, first edition.

Fare, R., Grosskopf, S., Lovell, C. A. K., and Pasurka, C. (1989). Multilateral productivity comparisons when some outputs are undesirable: A nonparametric approach. *Review of Economics and Statistics*, 71(1):90–98.

Fare, R., Grosskopf, S., Norris, M., and Zhang, Z. (1994b). Productivity growth, technical progress, and efficiency change in industrialized countries. *American Economic Review*, 84(1):66–83.

Farrell, M. J. (1957). The measurement of productive efficiency. *Journal of the Royal Statistical Society*, 120(3):253–281.

Fried, H. O., Lovell, C. A. K., and Schmidt, S. S., editors (2008). *The Measurement of Productive Efficiency and Productivit Growth*. Oxford University Press, USA.

Golany, B. and Roll, Y. (1989). An application procedure for DEA. *Omega-International Journal of Management Science*, 17(3):237–250.

Greene, W. H. (1980). Maximum likelihood estimation of econometric frontier functions. *Journal of Econometrics*, 13(1):27–56.

Hatefi, S. and Torabi, S. (2010). A common weight MCDA-DEA approach to construct composite indicators. *Ecological Economics*, 70(1):114–120.

He, F., Zhang, Q., Lei, J., Fu, W., and Xu, X. (2013). Energy efficiency and productivity change of china's iron and steel industry: Accounting for undesirable outputs. *Energy Policy*, 54:204–213.

Horta, I. M., Camanho, A. S., and Da Costa, J. M. (2010). Performance assessment of construction companies integrating key performance indicators and data envelopment analysis. *Journal of construction engineering and management*, 136(5):581–594.

Khalili, M., Camanho, A. S., Portela, M. C. A. S., and Alirezaee, M. R. (2010). The measurement of relative efficiency using data envelopment analysis with assurance regions that link inputs and outputs. *European Journal of Operational Research*, 203(3):761–770.

Koopmans, T. (1951). Analysis of production as an efficient combination of activities. In *Activity Analysis of Production and Allocation*, pages 33–97. John Wiley and Sons, New York.

Krautzberger, L. and Wetzel, H. (2012). Transport and CO2: productivity growth and carbon dioxide emissions in the european commercial transport industry. *Environmental & Resource Economics*, 53(3):435–454.

Kuosmanen, T. and Johnson, A. L. (2010). Data envelopment analysis as nonparametric least-squares regression. *Operations Research*, 58(1):149–160.

Kuosmanen, T. and Kortelainen, M. (2012). Stochastic non-smooth envelopment of data: semi-parametric frontier estimation subject to shape constraints. *Journal of Productivity Analysis*, 38(1):11–28.

Kuosmanen, T. and Matin, R. K. (2011). Duality of weakly disposable technology. *Omega*, 39(5):504–512.

Lins, M. E., Oliveira, L. B., da Silva, A. C. M., Rosa, L. P., and Pereira Jr., A. O. (2012). Performance assessment of alternative energy resources in brazilian power sector using data envelopment analysis. *Renewable and Sustainable Energy Reviews*, 16(1):898–903.

Liu, W. B., Meng, W., Li, X. X., and Zhang, D. Q. (2009). DEA models with undesirable inputs and outputs. *Annals of Operations Research*, 173(1):177–194.

Lovell, C. A. K. (1995). Measuring the macroeconomic performance of the taiwanese economy. *International Journal of Production Economics*, 39(1-2):165–178.

Lovell, C. A. K., Pastor, J. T., and Turner, J. A. (1995). Measuring macroeconomic performance in the OECD: a comparison of european and non-european countries. *European Journal of Operational Research*, 87(3):507–518.

Luenberger, D. (1992a). Benefit functions and duality. *Journal of Mathematical Economics*, 21(5):461–481.

Luenberger, D. (1992b). New optimality principles for economic-efficiency and equilibrium. *Journal of Optimization Theory and Applications*, 75(2):221–264.

Mahlberg, B. and Obersteiner, M. (2001). Remeasuring the HDI by data envelopment analysis. Laxenburg, austria, International Institute for Applied Systems Analysis, Laxenburg, Austria.

Managi, S. (2003). Luenberger and malmquist productivity indices in japan, 1955-1995. *Applied Economics Letters*, 10(9):581–584.

Meeusen, W. and van den Broeck, J. (1977). Efficiency estimation from cobb-douglas production functions with composed error. *International Economic Review*, 18(2):435–444.

Merriam-Webster (2013). Livability. in http://www.merriam-webster.com (19 july 2013).

Morais, P. and Camanho, A. S. (2011). Evaluation of performance of european cities with the aim to promote quality of life improvements. *Omega*, 39(4):398–409.

Nardo, M., Saisana, M., Saltelli, A., Tarantola, S., Hoffmann, A., and Giovannini, E. (2008). *Handbook on Constructing Composite Indicators: Methodology and user guide.* Organisation for Economic Co-operation and Development.

Newman, P. W. G. (1999). Sustainability and cities: extending the metabolism model. *Landscape and Urban Planning*, 44(4):219–226.

Newman, P. W. G., Birrell, B., Holmes, D., Mathers, C., Newton, P., Oakley, G., OConnor, A., Walker, B., Spessa, A., and Tait, D. (1996). Human settlements: Australia state of the environmental 1996. Technical report, CSIRO Publishing, Collingwood, Australia.

Newton, P., Baum, S., Bhatia, K., Brown, S., Cameron, S., Foran, B., Grant, T., Mak, S., Memmott, P., Mitchell, G., Neate, K., Smith, N., Stimson, R., Pears, A., Tucker, S., and Yencken, E. (2001). Human settlements: Australia state of the environment 2001. Technical report, CSIRO Publishing on behalf of the Department of the Environment and Heritage, Canberra.

Oggioni, G., Riccardi, R., and Toninelli, R. (2011). Eco-efficiency of the world cement industry: A data envelopment analysis. *Energy Policy*, 39(5):2842–2854.

Oregon Department of Transportation (2011). Livability and quality of life indicators. Technical report, Oregon Department of Transportation, in http://www.oregon.gov/ODOT/TD/TP/docs/LCP/Livability.pdf?ga=t.

Pichardo-Muiz, A. (2011). The role of diseconomies of transportation and public safety problems in the measurement of urban quality of life. *Applied Research in Quality of Life*, 6(4):363–386.

Pittman, R. W. (1983). Multilateral productivity comparisons with undesirable outputs. *Economic Journal*, 93(372):883–91.

Podinovski, V. V. and Athanassopoulos, A. D. (1998). Assessing the relative efficiency of decision making units using DEA models with weight restrictions. *Journal of the Operational Research Society*, 49(5):500–508.

Podinovski, V. V. and Bouzdine-Chameeva, T. (2011). The impossibility of convex constant returns-to-scale production technologies with exogenously fixed factors. *European Journal of Operational Research*, 213(1):119–123.

Portela, M. C. A. S. and Thanassoulis, E. (2006). Zero weights and non-zero slacks: Different solutions to the same problem. *Annals of Operations Research*, 145:129–147.

Ramon, N., Ruiz, J. L., and Sirvent, I. (2011). Reducing differences between profiles of weights: A peer-restricted cross-efficiency evaluation. *Omega*, 39(6):634–641.

Richmond, J. (1974). Estimating the efficiency of production. *International Economic Review*, 15(2):515–521.

Rogge, N. (2012). Undesirable specialization in the construction of composite policy indicators: The environmental performance index. *Ecological Indicators*, 23:143–154.

Sahoo, B. K. and Acharya, D. (2010). An alternative approach to monetary aggregation in DEA. *European Journal of Operational Research*, 204(3):672–682.

Sarrico, C. S. and Dyson, R. G. (2004). Restricting virtual weights in data envelopment analysis. *European Journal of Operational Research*, 159(1):17–34.

Scheel, H. (2001). Undesirable outputs in efficiency valuations. *European Journal of Operational Research*, 132(2):400–410.

Seiford, L. M. and Zhu, J. (2002). Modeling undesirable factors in efficiency evaluation. *European Journal of Operational Research*, 142(1):16–20.

Shephard, R. W. (1970). *Theory of cost and production functions*. Princeton University Press, Princeton, NJ, USA.

Shephard, R. W. and Fare, R. (1974). The law of diminishing returns. *Zeitschrift fr Nationalkonomie*, 34(1-2):69–90.

Simar, L. and Wilson, P. W. (1998). Sensitivity analysis of efficiency scores: How to bootstrap in nonparametric frontier models. *Management Science*, 44(1):49–61.

Simar, L. and Wilson, P. W. (2002). Non-parametric tests of returns to scale. *European Journal of Operational Research*, 139(1):115–132.

Thanassoulis, E., Portela, M. C., and Allen, R. (2004). Incorporating value judgments in DEA. In *Handbook on Data Envelopment Analysis*. Kluwer Academic Publishers, Boston.

Thompson, R. G., Langemeier, L. N., Lee, C. T., Lee, E., and Thrall, R. M. (1990). The role of multiplier bounds in efficiency analysis with application to kansas farming. *Journal of Econometrics*, 46(1-2):93–108.

Thompson, R. G., Singleton, F. D., Thrall, R. M., and Smith, B. A. (1986). Comparative site evaluations for locating a high-energy physics lab in texas. *Interfaces*, 16(6):35–49.

Thompson, R. G. and Thrall, R. M. (1994). Polyhedral assurance regions with linked constraints. In *New Directions in Computational Economics*, pages 121–133. Kluwer Academic Publishers, Dordrecht.

Timmer, C. P. (1971). Using a probabilistic frontier production function to measure technical efficiency. *Journal of Political Economy*, 79(4):776–794.

United Nations (1987). Report of the world commission on environment and development. Technical report, United Nations.

United Nations (1998). Kyoto protocol to the united nations framework convention on climate change. Technical report, United Nations.

United Nations (2013). 2013 human development report. Technical report, United Nations Development Programme, New York.

United Nations Development Program (2000). Human development report 2000. Technical report, Oxford University Press.

Williams, J., Peypoch, N., and Barros, C. P. (2011). The luenberger indicator and productivity growth: a note on the european savings banks sector. *Applied Economics*, 43(6):747–755.

Winsten, C. (1957). Discussion on mr. farrell's paper. *Journal of the Royal Statistical Society*, Series A(120):282–284.

Wong, Y. H. B. and Beasley, J. E. (1990). Restricting weight flexibility in data envelopment analysis. *The Journal of the Operational Research Society*, 41(9):829–835.

World Bank (2008). Environmental sustainability: An evaluation of world bank group support. Technical report, World Bank, Washington.

Zaim, O., Fare, R., and Grosskopf, S. (2001). An economic approach to achievement and improvement indexes. *Social Indicators Research*, 56(1):91–118.

Zhang, C., Liu, H., Bressers, H. T. A., and Buchanan, K. S. (2011). Productivity growth and environmental regulations - accounting for undesirable outputs: Analysis of china's thirty provincial regions using the malmquist-luenberger index. *Ecological Economics*, 70(12):2369–2379.

General index

Author index